NATIONAL QUALITY FORUM
+ + + + +
NEUROLOGY ENDORSEMENT MAINTENANCE
STEERING COMMITTEE
+ + + + +
THURSDAY
JUNE 21, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, NW., Washington, D.C., at 8:00 a.m., David Knowlton and David Tirschwell, Co-Chairs, presiding.

PRESENT:
DAVID KNOWLTON, M.A., Co-Chair, New Jersey Health Care Quality Institute
DAVID TIRSCHWELL, M.D., M.Sc., Co-Chair, University of Washington
A.M. BARRETT, M.D., Stroke Rehabilitation Research, Kessler Foundation

WILLIAM BARSAN, M.D., University of Michigan Health System
JOCELYN BAUTISTA, M.D., MBA, Cleveland Clinic
RAMON BAUTISTA, M.D., MBA, University of Florida HSC/Jacksonville
GWENDOLEN BUHR, M.D., Duke University

GAIL AUSTIN COONEY, M.D., FAAHPM, Hospice of Palm Beach County/Spectrum Health Inc.
JORDAN EISENSTOCK, M.D., CPE, UMass Memorial Medical Center
RISHA GIDWANI, Dr.PH., Stanford University Medical Center
DAVID HACKNEY, M.D., Beth Israel Deaconess

Medical Center
GREGORY KAPINOS, M.D., MS, North Shore-LIJ Health System

PRESENT(Cont'd):
MICHAEL KAPLITT, M.D., Ph.D., Weill Cornell Medical College
DANIEL LABOVITZ, Montefiore Medical Center
THERESE RICHMOND, Ph.D., CRNP, FAAN, University of Pennsylvania, School of Medicine
JACK SCARIANO, M.D., PLLC, Neurologist, Fort Sanders Parkwest Medical Center and Tennova West Medical Center
RAJ SHETH, M.D., Nemours Children's Clinic
JOLYNN SUKO, MPH, Virginia Mason Medical Center
JANE SULLIVAN, PT, DHS, MS, Northwestern University Feinberg School of Medicine
FREDRIK TOLIN, M.D., MBA, FACS, Humana
MARY VAN de KAMP, CCC-SLP, RehabCare, Kindred Healthcare
SALINA WADDY, M.D., National Institutes of Health
NQF STAFF:
HELEN BURSTIN, M.D., Senior Vice President, Performance Measures
ANN HAMMERSMITH, J.D., General Counsel
KAREN JOHNSON, MS, Senior Director, Performance Measures

KAREN PACE, RN, Ph.D., Senior Director, Performance Measures
SUZANNE THEBERGE, MPH, Project Manager, Performance Measures
JESSICA WEBER, MPH, Project Analyst, Performance Measures

ALSO PRESENT:
MARK ANTMAN, DDS, MBA, American Medical Association
SUSANNAH BERNHEIM, M.D., Yale New Haven Health System Center for Outcomes Research and Evaluation
JOHN BOTT, Agency for Healthcare Research and Quality*
ELIZABETH DRYE, M.D., SM, Yale New Haven Health System Center for Outcomes Research and Evaluation
JEFFREY J. GEPPERT, M.Ed., J.D., Battelle Memorial Institute*
NORBERT GOLDFIELD, M.D., 3M Health Information Systems*
KENDRA HANLEY, MS, AMA-PCPI
JEPH HERRIN, Ph.D. Yale New Haven Health System Center for Outcomes Research and Evaluation
DIEDRA JOSEPH, AMA-PCPI*
IRENE KATZAN, M.D., MS, Cleveland Clinic
KAREN KOLBUSZ, The Joint Commission
HARLAN KRUMHOLZ, M.D., Yale New Haven Health System Center for Outcomes Research and Evaluation*
JUDITH LICHTMAN, Yale New Haven Health System Center for Outcomes Research and Evaluation
ROB MULLEN, Ph.D., American Speech-Language-Hearing Association*
PATRICK ROMANO, University of California Davis Medical Group
SARAH TONN, MPH, American Academy of Neurology

ANN WATT, MBA, The Joint Commission
LAURA YODICE, MPH, MHA, American Medical Association
PAT ZRELAK, Ph.D., RN, University of California Davis Center for Healthcare Policy and Research *

*present by teleconference

CONTENTS

P-R-O-C-E-E-D-I-N-G-S

8:06 a.m.

MS. JOHNSON: Okay, good morning everyone. Thank you again for participating in our meeting. I know that we had a really great meeting yesterday, we got a lot accomplished.

So what we're going to do this morning, first thing we're going to change the agenda just a little bit. We're going to hand it over for a couple of minutes to David and Dave who are going to give us a quick recap of what happened yesterday.

And then we're going to hear from Karen Pace. Karen is another senior director here at NQF in the performance measures department and she is also our chief methodologist. And she is going to give us some background information just to help us think through some of the issues related to risk adjustment. So it will be a nice overview before we delve deep into the

mortality and readmission measures.

So with that I'm going to hand it over to our co-chairs.

CO-CHAIR TIRSCHWELL: Good morning, everyone. Welcome back.

So, the summary can be really quick. We reviewed I'm told 18 measures yesterday and 5 of them did not meet criteria for approval. Suzanne, can you tell us what those five were off the top of your head?

MS. THEBERGE: Yes, I can. It was the 0242 t-PA Considered, 2022 t-PA Initiated, 2017 CT or MRI Reports, and 0440 Stroke Education and 1955 NIH Stroke Scale Reporting.

CO-CHAIR TIRSCHWELL: Thank you. And today obviously we have a bunch more measures. It's a little bit different focus. A lot of these are outcome measures today as opposed to process yesterday. We're going to hear a little bit about methods which I think are inherently a lot more complicated today than they were yesterday.

And the schedule's been modified just a little bit. I'm not sure, when are we going to do the competing? At the end of the day. Okay, very good. Thank you.

DR. PACE: Good morning, everyone. I'm glad to be here. And we thought it would be good because the day is primarily devoted to outcome measures to give a little background and NQF perspective on risk adjustment.

These are, as has already been stated, more complex than the process measures that you've been looking at. It's -- we appreciate the questions and issues that people have been raising because it means you're taking a close look at the measures and identifying things and trying to understand what's going on. So we just want to give you a little background. And Jessica, you want to move to the next slide?

So, just to give you a quick background we've endorsed measures with a

variety of risk adjustment approaches. So, you're going to see three different approaches today in the measures you're looking at and there are more than that. And just to -- so the other thing from that is that the NQF criteria address risk adjustment, and I'll go into more detail about that, but do not dictate a specific statistical approach.

We don't require and we currently don't have a mechanism for head-to-head comparisons of different statistical risk adjustment approaches to the same data set. So it is complex, it's hard for, you know, for anyone to kind of get their heads around this when you start looking at different models.

Just, again, I think you're already well aware of this but NQF endorses performance measures for accountability applications and public reporting in addition to improvement, the primary goal. But our criteria apply to all applications. So currently we don't have different criteria for

different use cases, for example.

And the other thing to keep in mind is that NQF endorsement of the measure includes all the specifications but not reporting formats or presentations, for example, on how it's displayed on the web page for example. So, just to give you a little context there. Next slide.

So in terms of our criteria about risk adjustment you all know that we have a criterion about validity. And one of the elements under measure validity is for outcome measures and other measures where it's appropriate that there's an evidence-based risk adjustment strategy. Typically this will be statistical risk models, but occasionally it'll be risk stratification and that's something you'll see later today.

It should be based on factors that influence the measured outcome but not factors related to disparities in care or the quality of care. They should be risk factors that are

present at the start of care and have demonstrated adequate discrimination and calibration.

Occasionally we do get outcome measures that are not risk-adjusted. In that case we'd want to see some rationale and data analysis that supports that it doesn't need to be risk-adjusted. Next slide.

So, you noticed that there were a couple of little notes associated with that criterion. These are the specific notes. Risk factors that influence outcomes, we prefer that they not be exclusions, that they're actually in the risk model. And then note 14 is that risk models should not obscure disparities in care for populations by including factors that are associated with differences or inequalities in care. Typically race and ethnicity are the ones most thought of, but also socioeconomic status. Occasionally gender is associated with disparities.

And the whole point is that we're trying to identify disparities and get rid of them. So if we fold them into risk models then it's hard to really know that we have disparities going on and that we can do something about them. So that's the NQF perspective at this point in time. Okay, next slide.

So, I'm just going to talk a little bit high-level about statistical approaches for risk adjustment. And in the literature there seems to be emerging consensus on the need to address the correlation of clustered observations such as patients within hospitals and also to stabilize estimates of the performance. This is sometimes called smoothing, sometimes called shrinkage, sometimes referred to as reliability adjustment. But this is particularly an issue with small numbers.

Hierarchical models are appropriate to address both of these issues.

However, even within hierarchical models there are a variety of approaches. Even occasionally you can have non-hierarchical models that can address some of these issues. But all of those have different assumptions, strengths, weaknesses and practical considerations. Coming back to, you know, NQF has not dictated a specific statistical approach.

And also, as you all have started to look at the documentation for these measures, I'm sure quickly saw that comparison of those methods is very challenging. Next slide.

So, I'm going to talk about the CMS and AHRQ measures because those both have statistical models. They both use hierarchical approaches, they're both sound approaches that are supported in peer-reviewed literature and I think you also are aware of a white paper on statistical issues for performance measures. That was commissioned

by CMS from the Committee of Presidents of Statistical Societies that specifically looked at the hierarchical approach that CMS has been taking with their mortality and readmission measures. Next slide.

MS. JOHNSON: Karen, if we can interrupt you just a second. Developers on the line, if you would please mute your line, please. We're hearing some feedback from your lines.

DR. PACE: Okay, so the other thing is that although CMS and AHRQ measures are based on hierarchical approaches you noticed as you were going through them that they look different. And so I thought it would be helpful to at least kind of identify similarities and differences.

So both of them are addressing correlation of clustered observations meaning the patients within hospitals. Both of them stabilized or smoothed the hospital rate based on hospital-specific information in

combination with the national average. They both used the national model that includes only the patient-level factors as a comparison in the denominator, and they both compute the score for their measure as a rate. Next slide.

So, where do we see the differences is the modeling approach. So in the CMS measure it's accomplished in one step in the random effect hierarchical model. In the AHRQ measure it's accomplished in two steps. First, they used generalized estimating equations for clustering and then they do a reliability adjustment for smoothing. So they're both addressing the same issues, just in different statistical approaches and stages. Next slide.

So, given that it would be extremely difficult for us to ask you as a steering committee to look at these two and say, oh you know, one's better or somebody should have done it differently. You know,

the question is what do we really ask the steering committees to evaluate. And there's many things that we need your expertise on. Certainly our clinically relevant risk factors that are associated with the outcome included in the models. Are the risk factors those things that are present at the start of care? We don't want risk factors that are identified during or after care. The risk factors should not include those variables that are associated with disparities.

In terms of the statistical model, obviously you know was the statistical method appropriate for the data. And were the model performance metrics adequate. We do ask the developers to provide information about how the model is performing.

And I'll just make a note in terms of risk stratification when we get to those measures this afternoon or later this morning I guess, you know, the items about the risk factors are also applicable to stratification

variables. Those should be things that are associated with different levels of risk. And for stratification does an analysis demonstrate the relationship of those stratification categories to the occurrence of the outcome. So we'll be looking at those later this morning. And then next slide.

And finally, when you start -- after you've gone through the measures individually and you actually start to look at related or competing measures, you know, NQF does prefer to endorse measures with the broadest applicability, if possible to identify the best measure from among competing measures and certainly harmonized measures. As I mentioned, there's currently no mechanism to compare results using different statistical approaches but some questions still remain for the steering committee to consider. Are multiple measures needed? You know, if inpatient mortality is a subset of 30-day mortality, for example, how do those two work

together? And if Medicare patients are a subset of all patients. So those are some questions for you to consider.

Secondly, are the specifications and definitions of the outcome, the risk factors, the target population and exclusions, are those things harmonized across the measures? And then, you know, certainly we can have a discussion with the developers of whether they've discussed their approaches to identify the potential for and path to achieving one measure or harmonized measures.

So with that I'm going to stop and I know we have a lot to get into. But I'll also just say as, you know, as you're going through this process and you have any suggestions for us about our criteria, guidance to steering committees or information that we should be requesting from developers we would love to hear that. This is certainly an area that's been difficult not just, you know, for this steering committee but every

steering committee that comes up with outcome measures. So I'll stop there.

MS. JOHNSON: Thank you very much for that, Karen. Does anyone have any questions real quickly for Karen before we delve into the measures?

OPERATOR: For a comment or to ask a question press * then the number 1 on your telephone keypad.

MS. JOHNSON: Operator, this is not for public comment right now.

MEMBER WADDY: So, I had first of all a question about the definition of disparities that you all are using. So at least with the agreement between NIH and AHRQ it includes rural versus urban disparities which obviously would be pretty important when you're trying to determine quality care across the country. So my understanding is that that's not included in your definition?

DR. PACE: We just gave examples. We didn't have -- fully specify every type of

disparity. And I think the ones we identified are the ones that typically get considered to be put in as risk factors versus the rural/urban because that's more at a higher level than the patient level.

But you know, certainly, and I think our disparities task force has probably addressed that, but we just gave some examples.

MEMBER WADDY: And then just really quickly, I still am not entirely clear how you handled disparities. Do you have an example within what we're doing?

DR. PACE: It's a good question. In terms of the risk models, generally we don't want to see those variables in the risk model unless there's good data and analysis and evidence to indicate that they should be in for a particular outcome or reason. So that would be the exception rather than the rule.

And in terms of whether it's

outcome measures or process measures we don't currently -- what we're doing right now is trying to identify measures that would be disparity-sensitive so that they can be reported to highlight these disparities. But it's not generally been a part of every single measure. So, Helen may want to elaborate on that a little bit because I know she's been involved with the disparities work.

DR. BURSTIN: I think that captures it and we talked a little bit about disparity sensitivity yesterday. So the idea would be if you put race or ethnicity, for example, into the risk model you then can't stratify by it. And so the idea would be instead to be able to see those differences and stratify it, yes.

MEMBER KAPINOS: Can you go back to your first slide when -- somewhere I read that you did not validate your own risk model? Somewhere it says like you do not have the data to --

DR. PACE: No. What I was saying is that NQF does not have a requirement first of all or a mechanism. For example, for us to take the, for example, the AHRQ and CMS risk adjustment models, apply those to one data set and come up with yes, this one's better than that one.

First of all, even if we had a common data set that we could run those models we would still -- and they came up with different results we would still have the question of how would you know which result is the better result. So what we're saying is that right now you have to look at the measures individually and the question really is did the developer use an appropriate and, you know, accepted method of doing risk adjustment.

But right now we don't have the capacity at NQF to say, you know, we don't have a data set for example that we could tell the developers you have to run your measures

and risk models on our data set so that we can see a comparison.

MEMBER KAPINOS: Not on your data set because you don't hold a data set of course.

DR. PACE: Right. But we ask every --

MEMBER KAPINOS: Why not asking them to validate it first before they submit?

DR. PACE: Yes, yes, every -- that's part of our criteria, that the risk model should be evidence-based and should demonstrate adequate discrimination and calibration. And that's why on the measure submission form we've asked them to provide that information to you. Okay?

MS. JOHNSON: Any other questions for Karen? Okay, if not let's go ahead and go into our meeting then. And I'm going to hand it over to Dave.

CO-CHAIR KNOWLTON: We're going to start with 0467. Therese?

MEMBER RICHMOND: This is a currently endorsed measure, the Acute Stroke Mortality Rate, which is -- the steward is AHRQ. And it's a risk-stratified outcome measure with data coming from administrative records.

It looks at the proportion or the percentage of hospital discharges with an in-hospital death among cases with a principal diagnosis of stroke, either ischemic or hemorrhagic, for patients 18 years and older.

There are three exclusions: transfer to another acute care hospital, MDC-14 which is pregnancy/childbirth/puerperium and missing key data, discharge, disposition, gender, age, quarter or year of principal diagnosis.

We had quite a juicy conversation about this measure. And is the -- I don't know if the developer is here, but we posed a lot of questions to the developer. And you'll see about a seven-plus page response to our

questions so thank you very much.

The impact, we'll start with impact.

CO-CHAIR KNOWLTON: Before you go on, I went out of order. I wanted to ask Dr. Romano if he wanted to comment on his measure before we started.

MEMBER RICHMOND: Sure.

DR. ROMANO: Yes, good morning. This is Patrick Romano. I'm a general internist and professor of medicine at UC Davis School of Medicine in Sacramento. And I'm here representing the Agency for Healthcare Research and Quality.

I think on the phone with me is John Bott from the Agency staff, Jeff Geppert from Battelle Memorial Institute that leads our analytic team, and Pat Zrelak from my team at UC Davis who's a neuroscience nurse.

So, this -- I think you've really summarized this measure. It is a currently endorsed measure. It's one of a family of

measures that look at risk-adjusted inpatient mortality for major medical conditions. There are also very similar NQF-endorsed measures for heart attack mortality, for pneumonia mortality, heart failure mortality.

So, this measure is designed for application with administrative hospital data sets. It's designed for use with both state data sets that state health data agencies have as well as data sets that hospitals may use internally or within hospital systems. So, with that, thank you. CO-CHAIR KNOWLTON: Any questions for Dr. Romano? Probably not yet. Okay, impact, if you would please, Therese.

MEMBER RICHMOND: Great. This was a criteria that actually our group -- we had a group of four, three of whom voted. It had agreement at either the high or moderate level. We know there's a lot of strokes. Mortality rate in the U.S. is about 17 percent with the greatest risk of death in the first 30 days. And in 2008 almost half of stroke

deaths occurred in the hospital. So our group rated this either as high or moderate.

CO-CHAIR KNOWLTON: Questions? Comments? Vote on impact?

MS. THEBERGE: We have 19 high and 2 moderate.

CO-CHAIR KNOWLTON: Okay.

MEMBER RICHMOND: I guess we go to evidence which we don't need to do for the process, like the process measures but they did need to make a link. And we agreed that they made the link between structure, process and outcome in terms of the outcome on mortality. So our group also said yes to that.

CO-CHAIR KNOWLTON: Questions? Okay, let's vote.

MS. THEBERGE: Twenty-one yes, zero no.

MEMBER RICHMOND: Okay. In terms of opportunity for improvement or performance gap we rated this either as high or moderate.

There was a variation in risk-adjusted rates ranging from 73 per 1,000 to 136 per 1,000 and there were variations across all categories looked at, whether it was region of country, type of hospital ownership, the teaching status, the size of the city or the number of beds in the hospital.

CO-CHAIR KNOWLTON: Questions or comments? Okay, let's vote.

MS. THEBERGE: We're at 19 high, 2 moderate.

CO-CHAIR KNOWLTON: Okay. On to reliability, scientific acceptability: reliability 2a.

MEMBER RICHMOND: Okay, here comes the seven-page response to our question. So this is the juice of the discussion.

In terms of specification and reliability two of us ranked this as high and one with insufficient evidence. They use a noise -- a signal-to-noise ratio of 0.776 which is very good. That is a weighted

average of reliability estimates across providers and showing variations. So, I know one of our group members had some questions. I don't know if they were answered to your satisfaction but reliability in general I think we thought was high, yes.

CO-CHAIR KNOWLTON: Questions or comments? Okay, we can vote on that.

MS. THEBERGE: We are at 17 high, 4 moderate.

CO-CHAIR KNOWLTON: Validity.

MEMBER RICHMOND: Okay, validity. In our telephone conversation one of us ranked this as low and two, insufficient evidence. Thus we posed a lot of questions to the developer and you saw both an updated form as well as a seven-page response. So there really are three things to look at here. One is the establishment of validity, the impact of threats and the risk adjustment. And I'll just say a little bit about each and then we can talk.

I will say that, and I was an insufficient evidence and they provided a lot of evidence that I'm much more comfortable with. This was looked at in terms of face validity with an expert panel but really the substance is a criterion validity. They established for both the denominator, are we picking up a stroke diagnosis, comparing the administrative data to a gold standard chart abstraction with very good sensitivity and specificity. And also they provided additional information on the numerator in terms of picking up stroke mortality in the hospital.

They tested their models also with the exclusion of transfer, with and without transfers from acute care hospitals and found no statistical difference. A lot of the questions really centered around risk adjustment and they used, I think the introduction was really helpful in terms of they do use a hierarchical model, logistic

regressions and GEE to deal with clustering, include covariates for gender and age, and then use a system that I think is a proprietary system but the logic is available of APR-DRGs which is an all-patient refined diagnosis-related groups that includes a severity measure that -- and a risk-for-mortality measure. So it really includes that in the modeling. And the severity measures is defined as the extent of physiological dysfunction or organ system loss or function.

We asked a lot of information on this and specifically what was -- how it was done and what was included in the model. And for the most part my questions were answered by that. They have a C statistics for the risk model and the development sample of 0.86 and the validation sample of 0.89.

I'm going to ask my other group members to jump in here because we had so many questions on the risk adjustment if I could.

CO-CHAIR KNOWLTON: Sure. Other

group members on this issue? Risha?

MEMBER GIDWANI: Good morning. I had a couple of different questions. I posed a number of them during the original work group call but the developers did a wonderful job actually of responding to those.

The ones that I have remaining I'd like the developers to confirm that the coefficients they're presenting are log odds. Is this correct?

MR. GEPPERT: Yes, that's correct. I'm sorry, this is Jeff Geppert from -- that's correct. Yes.

MEMBER GIDWANI: Okay. A recommendation for the future is to actually present these in terms of probabilities. Log odds are actually quite difficult to interpret, so to either exponentiate them as odds or to use actual probabilities.

In terms of the APR-DRGs there's a risk of mortality and that's the last digit in that four-digit number under "Label." And I'd

also request the developers to confirm that these APR-DRG risk of mortalities are actually for the admission status rather than the discharge.

DR. ROMANO: Yes, that's correct. They're based on the diagnoses that were present on admission.

MEMBER GIDWANI: Okay, great. Thank you. And then just a couple of other things.

One, the APR-DRGs are a black box. 3M owns this methodology and they do not provide the details of how risk of mortalities are calculated to anybody. So a suggestion is to move in the future away from a methodology that has a black box associated with it and to go towards something more transparent.

And then finally I would like to ask the developers whether there is a rationale for -- or if there's a thought towards excluding EMTALA patients who may be at higher risk of mortality yet the hospitals

are not able to turn these patients away. They're coming in through the ED. And whether this would actually unfairly ding hospitals that see a lot of EMTALA patients.

CO-CHAIR KNOWLTON: Dr. Romano?

DR. ROMANO: I'm sorry, how would you identify or define EMTALA patients?

MEMBER GIDWANI: Well, I'm not sure if there's a billing code for that, but there is no status of whether the patient came in through the emergency department and what that patient's level of severity was when coming in through the ED. So there's not even an ability to understand whether they came in through Life Flight or ED which would render them at higher risk of mortality.

DR. ZRELAK: This is Pat Zrelak from the UC Davis team. The majority of your stroke patients will come in through your emergency department.

DR. ROMANO: Right, the greater majority are ED admissions. I'm afraid

there's no data element that would specifically distinguish those who might be classified as EMTALA patients or those who would be brought in by helicopter.

With reference to your first comment, I believe that Dr. Goldfield is on the line but there's actually a limited license agreement between AHRQ and 3M that effectively puts the components of the APR-DRG system that are necessary for risk adjustment to this indicator into the public domain. And I'll ask Dr. Goldfield or Dr. Geppert to comment on that further.

DR. GOLDFIELD: I'm on the line. Maybe Rich Averill who's also on the line could say specifically exactly on that point. So I appreciate being asked because in fact I would have to differ with the assertion that it's a black box.

MR. AVERILL: We have a website that's APR sign with a login. And anybody who wants to inspect the complete APR risk of

mortality logic can request a copy of that and at no cost can fully inspect all aspects of the logic. We encourage people to do that. We solicit comments and that is part of our annual update process.

DR. GOLDFIELD: And I just would like to add -- this is Norbert Goldfield speaking -- just briefly, I don't want to take too much time, is that the critical aspect which is why the APR-DRGs are extremely widely used is the fact that it's a categorical model and clinicians drill down, right down to the individual patient level. And that's also why we encourage --

CO-CHAIR KNOWLTON: Would you speak up please on the phone? It's hard to hear you in the meeting room. Just speak up a little.

DR. GOLDFIELD: Can people hear me now?

CO-CHAIR KNOWLTON: Yes, that's much better. Thank you.

DR. GOLDFIELD: I was just making the point that the APR-DRGs are a categorical clinical model which means that similar but a different model to the MS-DRGs but applied to all patients. The clinicians can drill down, right down to the individual patient and see exactly for that patient why the person was assigned to a particular severity level which is why we encourage strongly individuals to not only access the model but provide feedback and sort of a consequence. It's not a black box, but appreciate the opportunity to comment.

CO-CHAIR KNOWLTON: Risha?

MEMBER GIDWANI: Thank you, I wasn't aware of that. I wasn't able to actually access information on how the APR-DRG risk of mortality was assigned so I'll look into that if that is available publicly.

I still, however, do have a concern. I think generally these models are doing a good job but if they are able to

account for Life Flight patients or EMTALA patients due to a lack of billing codes I see that that, you know, poses a logistical difficulty but in terms of actually accounting for risk of mortality the models wouldn't be able to do that. So it's one component that is not present here.

CO-CHAIR KNOWLTON: Can I ask you a question, Risha? When you're talking about an EMTALA patient you're not just talking about insurance status. That's really the issue, is that not true?

MEMBER GIDWANI: Well, these are patients that are coming into the emergency department that cannot be turned away because they are, you know, having a true emergency.

CO-CHAIR KNOWLTON: But the basis for turning them away would be that they do not have insurance or in some way cannot afford the care. Because other than that as I think has been pointed out all of them, or not all of them but a majority of them are

coming in through the ER.

So what differentiates an EMTALA from a non-EMTALA is merely the ability to pay. That would be the only reason that you would activate EMTALA would be that you have to treat them under the EMTALA law because they don't have the ability to pay.

MEMBER GIDWANI: Right and --

CO-CHAIR KNOWLTON: And I wonder back to our risk stratification issue that Karen talked about, if we would be -- want to apply a risk factor in advance that has to do with that factor.

MEMBER GIDWANI: I think that's an interesting point. I think there's also other risk factors associated with lack of insurance that may be putting these patients at higher risk of mortality.

CO-CHAIR KNOWLTON: Okay. David and then Bill.

CO-CHAIR TIRSCHWELL: So I had a question for the developers and maybe you can

just confirm this. I think you already said the severity measures are all based on present-on-admission characteristics. But I mean are there symptom-specific present-on-admission characteristics like coma or something like that that gets you severity? And I see Dr. Romano is nodding his head. So I'll take that as a yes. And specifically complications are not included, things like pneumonia and things like that. He's nodding yes again.

But then the part two of my question, I know that you've checked for reliability as to identification of patients and such, but at the end of the day your models theoretically allow you to rank hospitals. And my question is have you taken this model which gives you a set of rankings and then compared a set of rankings to a gold standard patient database where there is detailed clinical information about patients. NIH Stroke Scale Scores, you know, chart-

abstracted comorbidities, a much -- obviously more expensive approach to predicting outcome but what I think most people would think of as a gold standard, and shown that the ratings are essentially highly correlated.

DR. ROMANO: We agree that that would be an important thing to do and in fact we've started having some discussions with the Get With the Guidelines group, American Heart Association group about -- they have as you probably know a linked data set with Medicare claims as well as registry data with detailed physiologic information and the NIH Stroke Scale. So we are hoping to use that information as kind of a laboratory for testing the comparative model.

They published a paper which you may have seen suggesting that the administrative data didn't perform nearly as well as models with NIH Stroke Scale. But the C statistic on the model that they were testing was in the range of 0.65 as I recall

whereas the C statistic on our model is close to 0.9. So it's a substantial gap in performance there. So we would hope that that would narrow the discrepancy in the kind of comparative analysis that you're describing but we haven't empirically tested that.

What we have empirically tested that I could comment on in follow-up to the comments. One is that there was some concern about the fact that the APR-DRG risk adjustment does incorporate some information about procedures that are performed during the hospital stay. And that could be construed as a violation of the NQF principles that Karen described.

So for example, if a patient has a hemorrhagic stroke and they require a craniotomy for evacuation of the hematoma then that craniotomy goes into the risk adjustment. And it is an important factor in the risk adjustment in terms of the likelihood function, but we've actually -- in follow-up

to the discussion we tested models without those variables and we showed that the C statistic is essentially unaffected, that it remains about 0.9 without those procedure-based APR-DRGs. And furthermore the correlation in a provider rate between models with and without those procedure-based APR-DRGs is 0.978.

So basically although those procedure factors are correlated with the expected mortality rate and with the observed mortality rate, they don't actually explain variation across hospitals. So they serve as proxies for stroke severity essentially. So having said that, they -- we could basically go either way in terms of including those factors or not.

The other thing that I should say that we tested in terms of a gold standard analysis is using the California state data set where we had the ability to look at 30-day mortality as well as inpatient mortality. We

re-estimated the risk adjustment model using the California data set, again demonstrated a very high C statistic of 0.863. And looking at the correlation, the weighted correlation of risk adjustment inpatient versus 30-day mortality at the hospital level, that correlation was 0.64. So that's in the moderate range.

CO-CHAIR TIRSCHWELL: That's r-squared or just r?

DR. ROMANO: That is r.

CO-CHAIR TIRSCHWELL: So then your r-squared is only 0.3 so you're only explaining about 40 percent of the variation in the 30-day mortality with your inpatient model which doesn't seem fantastic to me.

DR. ROMANO: Correct. So there is some difference between inpatient and 30-day mortality measures. What we and others have demonstrated is that there's more hospital-level signal. So if you look at what we call the intra-class correlation coefficient, the

hospital-level signal there's more signal looking at inpatient mortality which of course makes sense because it's more a reflection of what happens in the hospital, less affected by what happens after the patient is discharged. But it does potentially introduce the possibility of bias related to variation in transfer practices and length of stay across hospitals.

CO-CHAIR TIRSCHWELL: Yes, I mean the objection that was raised was that hospitals whose practice is to discharge all comfort care patients to a nursing home to die will artificially look like they're performing better, whereas the 30-day mortality measure would even that out theoretically.

And you know, part of my impression of the big difference between your model and for example the one you referred to in Get With the Guidelines, and I'm not 100 percent sure about this, is that they limited their analysis to ischemic strokes. And by

having the three subtypes of stroke in your model my guess is that the vast majority of your explanatory power is based on the fact that you can separate hemorrhages with their much higher mortality rates from ischemic strokes with a much lower mortality rate.

And I think you'd find shocking differences in your C statistics if you stratified your model by stroke type. In fact, I think your performance would radically fall in your C statistic if you looked at the model for each stroke type separately.

So, you know, I think this is an amazing thing to do and have out there. The fact that God knows how much money AHRQ has spent on this measure already and the fact that you haven't spent, you know, probably a relatively small amount of money to validate it against a high-quality, carefully abstracted set of patient-specific data is a little disappointing.

CO-CHAIR KNOWLTON: Michael?

MEMBER KAPLITT: So, you know, along the lines of I guess this or the EMTALA question. I was looking at various stratifications and unless I'm not seeing it why is there no discussion of transfers as an issue like either in or out? Because along the same lines -- so if you're a hospital that transfers out a lot of patients with hemorrhages to my hospital, we operated on them, they die. Our hospital looks like we don't do very well with strokes, your hospital looks like you're great and then everybody wants to go there but that's because you're transferring out all the people that are sick to my hospital. Right? So how is that accounted for and why isn't that? Or is it and I'm not seeing it?

DR. ROMANO: We do routinely test for transfer in as a risk factor for mortality. And we find that in some patient cohorts it's a significant risk factor and in others it's not. In this particular model

actually it didn't enter the model. I'm not sure, Jeff, do you have any additional comment on that?

MEMBER KAPLITT: Before you answer I would also ask about transfers out because I would think that's equally relevant. You know, if somebody's actually doing better because they're transferring a disproportionate number of patients I would think that should be in the model.

DR. ROMANO: Well, remember that this measure is designed for hospitals themselves to use, hospital systems, state health data agencies, regional coalitions, other entities that don't have the ability to link data across hospitals. So, therefore the patients who are transferred out are excluded from the analysis because we don't know what their outcome status is at the time that they leave the acute care center. So those patients are excluded from both the numerator and the denominator.

MEMBER KAPLITT: That I understand. I'm just saying that like later we're going to be asked in one of the later sections about the potential for misuse, let's say, or problems or whatever in one of the later things and we know that there are plenty of areas. If you look at the history of let's say cardiac surgery testing or whatever, when these things get implemented there are some incentives for certain institutions to, you know, not treat patients that are sick or transfer them out. And I would think that in order to disincentivize that, that should be somehow included in the model, you know, equally to transfers in.

MR. GEPPERT: We have in fact tested hospital-level models where we used transfer-out percentage as a factor in the model and it doesn't have explanatory power in those hospital-level models relative to other things, other characteristics of the hospital like their volume or their capacity.

CO-CHAIR KNOWLTON: Salina.

MEMBER WADDY: I agree with Michael's point. That's exactly what we've been sitting here discussing.

I do think it would be very interesting to know down the line if when this is implemented for -- or if this is implemented through this as well as in the past few years that they've had the original endorsement whether or not that actually changes what hospitals potentially do in terms of how they handle the -- handle the transfers.

The other thing is what happens to the patients. And this happened frequently when I was at Emory. Patients that were transferred out but they did not end up being admitted at the other hospital, at the receiving hospital because of death en route.

CO-CHAIR KNOWLTON: Ramon?

MEMBER R. BAUTISTA: This is a very important measure. In fact, it probably

is the sole measurement that might make or break many stroke centers.

It's also a reevaluation of a measure we actually saw in 2008. So similar to the stroke education measure from yesterday I think the committee should demand evidence of it being used properly out there before approving it for re-implementation again. Otherwise we're unfairly penalizing some stroke programs and that would be very bad for them.

CO-CHAIR KNOWLTON: Daniel?

MEMBER LABOVITZ: I think this is a corollary to what Ramon just said. But I think death as a measure of quality is a really complicated area. I'm not sure -- death is not a common stroke outcome. Stroke is I guess now the fourth leading cause of death, but those deaths don't occur in the hospital.

I think what hospitals do with patients who are near the end of life or at

the end of life can range widely in quality. And I've worked in a few different New York City area hospitals. One of the hospitals that does the best on death does the worst on compassion.

I think we could really drive -- we could really go the wrong way here with a measure that is based on administrative data. And I would be very, very reluctant to put a stamp of approval on that.

CO-CHAIR KNOWLTON: Bill?

MEMBER BARSAN: I just have a question for the developer again. So the question was asked about the transfers in and I think you -- what I took from that is that you use the transfer things for different measures but not for this one? Is it used for this one or not?

DR. ROMANO: It was not statistically or clinically significant in the model for this measure so it was excluded from the model.

MEMBER BARSAN: So you don't use it.

DR. ROMANO: Correct.

MEMBER BARSAN: Okay.

CO-CHAIR KNOWLTON: Risha?

MEMBER GIDWANI: I also have a question about patients who are on DNR or DNI status. And I understand that that's not accounted for in this model either, is that correct?

DR. ROMANO: That's correct. There is a data element that's been introduced in some states and we're doing some exploratory testing using that data element but there's obviously some concern about when the order is written, whether the order is written after some deterioration of the care of the patient as well as variation across hospitals.

So again, getting back to the methodologic concerns that Karen raised, we want to make sure that that in itself is not

a quality issue before we include it in the risk adjustment model.

MEMBER GIDWANI: I'm not a clinician so correct me if I'm wrong, but isn't it the patient and/or family decision to be a DNR or DNI?

DR. GOLDFIELD: Could I just comment on that? This is Norbert Goldfield. I just want to say in addition there's a lot of literature on this point. In fact, one reason to be very careful about the DNR is because of practice pattern variation. When Patrick was referring to the hospital the hospital is clearly composed obviously of patients and their families and physicians, and there's a lot of practice pattern variation in terms of preferences. And one just has to be very careful about incorporating that into the risk model. All right, Patrick.

MEMBER GIDWANI: But what we're talking about incorporating is not physician

preference, we're talking about incorporating patient and family preference.

DR. GOLDFIELD: To put it simply that information is not available today. A and B as you probably know, again the literature is quite clear on that point too. It's not -- that's not done in a vacuum. That very much can be impacted by provider preference or professional preference.

CO-CHAIR KNOWLTON: Are you done, Risha?

MEMBER GIDWANI: Yes.

DR. ROMANO: And it's also been documented that in many cases the DNR orders are written after some events happen in the hospital. So it may in fact be a marker for deterioration of the patient after admission to the hospital.

MEMBER KAPINOS: Actually and to clarify about that, it shouldn't be only looking at the DNR/DNI because actually palliative care specialists and the people who

work in ICUs have tried to -- recently tried to separate the whole concept of do not resuscitate as opposed to goals of care. So we shouldn't be talking about DNR because DNR/DNI just means do not resuscitate upon a cardiac or a respiratory arrest as opposed to what we are all discussing is the decisions of how aggressive should be the level of -- how aggressive should be the goals of care. So we should reformulate this discussion about not DNR/DNI but actually the level of aggressivity of the goals of care. That's actually a more proper terminology.

And to answer your question, actually no, it's not -- I mean a lot of ethicists, I mean Bernard, you know, like the famous -- the trial on the cardiac arrest recently also published in Neurology, the fact that actually the more you offer to patient family members the opportunity to consent or to give their opinion on should the patient be resuscitated or not, the more guilt you

trigger in those family members.

And actually more -- many palliative care physicians are actually saying that DNR/DNI is based more on if you think as a physician that the patient should not be resuscitated you should not ask a question to the family members, but you tell them that you think it does not make sense to fight for such a low quality of life.

So I just want to make sure that people don't look at the issue of DNR/DNI as a black and white thing. I think it's more goals of care and it's more gray with many shades as opposed to just something that's going to be easily measurable.

CO-CHAIR KNOWLTON: Jolynn?

MEMBER SUKO: Yes, I just wanted to present a counterpoint to Daniel's argument around outcome. And I think that is that yes, there's many variables that lead to death but if we're not looking at outcome we're not going to be able to identify those

institutional level variables that may contribute to quality and how we organize care.

We know in many situations for clinical conditions there's a relationship between volume and outcome. And I think that there's no performance measure that's going to be perfect. And so yes, there's a risk of misuse but there's also a risk of not looking and not continuing to draw on those questions by not endorsing an outcome measure.

CO-CHAIR KNOWLTON: Gail?

MEMBER COONEY: I think I just retracted my comment.

CO-CHAIR KNOWLTON: Okay. David?

MEMBER HACKNEY: I'm going along with Jolynn on the importance of looking at outcome measures, but what I'm worried about is it's going to be almost impossible to interpret this one. You'll get data but you won't know what it means and it could be highly misleading and point you in the wrong

direction about which hospitals and practices are doing well and which are doing poorly. So I'm -- without knowing what the results mean I don't know how we can endorse a measure.

CO-CHAIR KNOWLTON: Michael?

MEMBER KAPLITT: So, you know, I'm sorry to belabor this point but I know you said, and that's a good answer, that you know, you looked at transfers and it wasn't significant. But as David said earlier you're looking at multiple different stroke types, not just ischemic stroke.

Did you look at transfer issue by stroke type? And the reason I ask that is that let's say one hospital has 1,000 ischemic strokes and 10 hemorrhagic strokes, and they transfer all 10 to my hospital. My hospital also has 1,000 ischemic strokes but I take all 10 hemorrhagic strokes because I have neurosurgeons and they don't. Nine of those ten die because they have the higher mortality rate so that my mortality rate's going to be

higher because of those 10 people but it's not going to show up statistically in the entire population because it's being washed out.

So, have you looked at it that way? Because I just don't want hospitals that have certain level of care that's actually providing better care to be penalized because they're taking what represent the majority of the deaths with the minority of the transfers.

DR. ROMANO: Okay, so I think you're suggesting an interaction or effect modification between the type of stroke and whether the patient is transferred in. Jeff?

MEMBER KAPLITT: Transfers of the type --

DR. ROMANO: To my knowledge we have not looked at that. The model does stratify risk of mortality separately by hemorrhagic versus ischemic stroke. So there is the opportunity for different risk factors to affect the risk of mortality for hemorrhagic versus ischemic stroke but that

does not apply. That applies to clinical risk factors such as the coma at presentation and so forth. Doesn't apply to transfer status.

Jeff, do you have anything to add to that?

MR. GEPPERT: Just to go back to the earlier comment about whether the explanatory power was due to the inclusion of both stroke types. And we did test that. We examined our C statistic separately for each stroke type.

Patrick, do you happen to have --

CO-CHAIR KNOWLTON: Let's not go back and forth on the measure. We're getting into the weeds deeper and deeper, so let's focus on what the committee has as questions. Risha?

MEMBER GIDWANI: I'm sorry, I think that's actually important. I would like to hear the C statistic that the developer --

CO-CHAIR KNOWLTON: Well, then ask the question again.

MEMBER GIDWANI: Okay, I'll just ask the developer. Can you please continue and present the C statistic for the models when you disaggregated by stroke type?

MR. GEPPERT: I was just going to ask Patrick if he happened to have those results with him handy. Otherwise I'll find them quickly.

But my recollection was that there was a slight drop in the C statistic but they were comparable. It was maybe like a 0.7 or 0.8, in that range, rather than a 0.9, something like that.

MEMBER GIDWANI: Well, a 0.7 is pretty different than a 0.9 so if we could actually get that statistics that would be great.

MR. GEPPERT: If I can't find it now we'll provide it to the committee later.

CO-CHAIR KNOWLTON: Do you know have another point while --

MEMBER GIDWANI: I do. I do. I

am actually quite comfortable with the C statistic of 0.9 that the model developers are presenting. But to get at this question of whether there are going to be some issues of model fit when we're looking at the extremes of the types of patients, did the model developers do a Hosmer-Lemeshow test? And if so, can you present those results?

MR. GEPPERT: We run that test because we have so much data it always rejects so it's not a particularly -- we don't find it to be a particularly useful diagnostic. So we tend to look at the risk decile charts.

DR. PACE: And that's consistent with what our -- some previous expert panels have told us about the Hosmer-Lemeshow statistic. And they did provide the risk decile plots for you in the response in terms of looking at the calibration of -- and that was in the responses from AHRQ.

CO-CHAIR TIRSCHWELL: It's in the Final Measures folder. And I think it's got

the --

DR. PACE: Suzanne, can you bring it up on the?

DR. ROMANO: Could I address one of the other questions?

DR. PACE: Yes, go ahead.

DR. ROMANO: So, I just wanted to say that we certainly agree with the importance of compassion and with the importance of other stroke measures. So this committee is considering and NQF has previously endorsed many other measures of stroke quality. This is certainly just one measure of what would be a comprehensive dashboard of measures related to stroke mortality which should certainly include measures related to patient experience and ideally functional outcomes as well.

This measure has been in use for 4 years with NQF endorsement. So I think you might look to the experience or lack thereof in terms of whether a measure has been misused

in leading to pernicious practices. We're not aware of that experience but we certainly would like to learn about it if committee members are aware of that kind of misuse of the measure.

CO-CHAIR KNOWLTON: Is there evidence of use for productive purposes? I guess I'm not aware of use in either direction quite honestly. From what Dr. Bautista was suggesting which seems reasonable 5 years hence, 4 years hence where's the evidence this is driving practice in a positive direction?

DR. BOTT: This is John Bott with AHRQ. And we did a series of user group meetings a couple of years back. We called it a Learning Network. And this measure was used as a case-in-point by a hospital coalition in Texas where they used this measure with the 70-plus hospitals in their association, found out what was driving mortality in some hospitals related to stroke and made improvements. I looked at the PowerPoint this

morning. They talked about it. They didn't have particular PowerPoints in that slide. That's at least one example.

DR. ZRELAK: This is Pat Zrelak again from the AHRQ team. And one of my other roles is to actually run the UC Davis stroke program. And so we do look at our stroke mortality and we're a hospital that has very high uninsured, very difficult stroke population. And we have a fairly high stroke mortality. And so I look at it right when I do my annual quality report, the hospital quality department, they want to hear about our stroke mortality and what we're doing to improve it. And when I do pull those cases and drill down there is a lot of opportunity there for improvement.

CO-CHAIR TIRSCHWELL: Are you calculating your AHRQ mortality ratio and putting it in perspective or are you just looking at your hospital stroke mortality cases which are obviously two totally

different things?

DR. ZRELAK: I do both. I use the AHRQ measure a lot for benchmarking so I can compare myself mainly against other university hospitals because I for the most part use the University Health Consortium measures. So I do both.

CO-CHAIR KNOWLTON: Ramon, then Risha.

MEMBER R. BAUTISTA: Let me just -- a counterpoint to the counterpoint. I mean, of course we would like to have a mortality measure. Of course we'd like to have a good education measure. But it has to be done well. And again, this is a reevaluation of an old measure. Unless done right I just feel that we will have negative consequences that's going to really hurt more people than help.

CO-CHAIR KNOWLTON: Risha?

MEMBER GIDWANI: I'm looking at the risk deciles that the developers provided and there are no values for the y axis so it's

difficult to understand what the difference is between observed and predicted.

CO-CHAIR TIRSCHWELL: It says mortality rate.

MEMBER GIDWANI: I can't see --

CO-CHAIR TIRSCHWELL: You don't know what the absolute values are.

MEMBER GIDWANI: Right, I just -- I don't know what that means. On the printout it shows. Okay.

And then the other thing is I'm looking at this same document. And in this document I also asked for an explanation of how x is an improved vector of binary explanatory variables compared with z. The response, if you scroll down you'll see it on this larger screen as well, the response is that x are covariates based on all secondary diagnosis codes while z are covariates based on secondary diagnosis codes that are coded as present on admission.

I'm unclear from this response,

I'm not sure if this was a misworking. Are you actually using covariates that are also not present on admission?

MR. GEPPERT: Just to address that question. So if the data has present-on-admission data elements on it and they are using coded present-on-admission data then no, we're not using covariates that do not use present-on-admission. If that's clear.

MEMBER GIDWANI: I'm sorry, I can't understand that. Are you using x as the model, as the predictor variables, or z as predictor variables?

MR. GEPPERT: We're using x which is the version of the predictor variables that uses the present-on-admission data.

MEMBER GIDWANI: That's not what's noted here in your response.

MR. GEPPERT: It might just be inverted, but x is the version that uses the present-on-admission data. Z is the version that does not. And we're using x.

MEMBER GIDWANI: Okay.

CO-CHAIR KNOWLTON: Helen?

DR. BURSTIN: I just want to make one point that we're still on validity. So many of the issues we've now stumbled into, they're really important, are usability. And I just want to make sure we keep our conversation separate. There's actually a great deal of detail under the usability section about current use. So.

MEMBER SUKO: Yes, I just wanted to speak to the use of outcome measures. And I can say that in the organization where I work we do actually look at our mortality. We go in and we'll do chart reviews on different clinical diagnosis. And we've actually -- we do. Oh, well I was responding to the outcome question.

CO-CHAIR KNOWLTON: Go ahead, finish.

MEMBER SUKO: Okay. So we do actually and we have discovered -- we've

discovered residents who had some opportunities for learning improvement in how they do things and how they document things. And so we have used outcome data. And we've also discovered documentation errors or things that were more administrative in nature.

And so while you can't say that it's always a clinical process that's gone wrong or an error in judgment by a provider you do discover things that lead to better management of patients.

CO-CHAIR KNOWLTON: Anybody have anything -- Therese, do you have anything else? We are on the issue of validity, okay? Everybody remember that from way back then? Are you ready? Let's vote. The issue is validity.

MS. THEBERGE: Two high, ten moderate, six low, four insufficient evidence.

CO-CHAIR KNOWLTON: Okay, we move on. Therese?

MEMBER RICHMOND: Whoa. Okay, now

we're up to usability. Our group ranked this, not surprisingly, one high, one low, and one insufficient evidence.

In terms of usability there are 18 states or systems that are said to publicly report this although not all of them when you go into the systems actually report stroke outcome, but many do. And it's also used in the Commonwealth Fund, Why Not the Best and Monarch. It's used for quality improvement by the University Health Consortium and in the Premier Quest tool. So we were across the board as a group.

CO-CHAIR KNOWLTON: Thoughts, comments? Jolynn, is your hand up? No. Risha?

MEMBER GIDWANI: I was the work group member that had a lot of questions, concerns and rated things oftentimes insufficient evidence. I will say that the developers did really an admirable job of responding to my questions and alleviating

many of my concerns. So when you see those values and I'm the left side outlier I would modify a lot of my conclusions now based off of feeling more comfortable with the models.

CO-CHAIR KNOWLTON: Salina?

MEMBER WADDY: So I was just wondering regarding the transfer issue -- yes, I'm back to that -- have they looked at if a patient is at hospital A and then is transferred to hospital B then it was that hospital's decision to transfer, or that physician or whatever, decision to transfer the patient. So can whatever happened to the hospital -- to the patient in hospital B actually be attached to hospital A? And whether or not that changes any of the appearance of the mortality at the hospital.

CO-CHAIR KNOWLTON: Dr. Romano?

MEMBER WADDY: Does that make sense?

DR. ROMANO: Yes, that is -- that is potentially possible with linked data sets.

And I think you'll hear shortly about the CMS measure that does precisely that.

It's not -- obviously the practical problem is that many users who are interested in looking at stroke mortality don't have access to linked data across hospitals. So we offer an alternative measure that's based on the hospital's own outcomes. But it is theoretically possible and it is done with the CMS measure.

CO-CHAIR KNOWLTON: Greg?

MEMBER KAPINOS: I just wanted to make a comment on -- I think it's about usability.

CO-CHAIR KNOWLTON: Lean into your mike. It's hard to hear you.

MEMBER KAPINOS: Sorry. So, I wanted to make a comment and I hope it's under this section of usability but earlier on when we were discussing about the fact that it's been implemented for 4 years and there's no proof of misuse. Then Dr. Tirschwell asked

well, what is the proof that it's actually used in a good way. I would -- then somebody also said something about the fact that while there's no -- there's no misuse and we can look into subcategories of -- or like linkage between mortality and other things. So it's going to be useful to have this measure.

I want to say that maybe in 2008 we were not that close from the government actually using those measures to make really like big decisions on how much money will get to each hospital as opposed to now we are I think very much closer. So there could be a misuse in the very near future about a mortality measure that is actually not valid and not capturing actually good quality of care.

And number two, I wanted to say is it not okay to not endorse one measure here? We're not throwing all the work of the AHRQ in the trash can, right? If it's not endorsed by NQF we still can collect that -- I mean, that

Agency will still collect that data and whoever is interested in actually using that model to calculate what should be their stroke mortality can still use that data.

So I'm just saying that my understanding is that NQF measures will be potentially used by the government as a standard and dinging hospitals that don't do a good job in terms of mortality. And I see an issue with that. And I'm just saying that maybe actually we can feel better about not endorsing some measures because actually there's also other agencies like the Joint Commission and AHRQ that will continue to do their job with those measures. Or am I wrong?

CO-CHAIR KNOWLTON: I wouldn't say that you're wrong but I wouldn't want to minimize the impact of NQF endorsement in terms of what the government accepts, what CMS accepts, what payers accept. It has great sway in terms of what happens. This is a consensus organization that has pretty deep

imprimaturs. So on the one hand what you say is true. On the other hand I wouldn't want to minimize the effect and be casual about it and say oh well, they can do it anyway because it has some real impact in the real world.

Other comments on this or are we ready for a vote? Therese, you all set? Oh, I'm sorry, Dr. Romano, you had another comment.

DR. ROMANO: Well, I just want to be clear and NQF staff can add to this maybe, but I think that this process is really about measures that can be used for transparency and accountability. There's really a separate process for measures that would be used for payment which has to do with what's called the Measure Applications Partnership. So, this endorsement I don't think implies that the measure would actually be used for hospital payment which is a very specific pay-for-performance application. So, just to be clear about.

DR. BURSTIN: So, NQF endorsement implies the measure is acceptable for a wide range of accountability applications, from certification all the way through. So again, the Measures Application Partnership which Patrick's referring to will make recommendations on specific measures to be used for specific programs, but that, you know, again the assumption should be these measures would be ready for all accountability applications.

CO-CHAIR KNOWLTON: Right. Yes, Dan, go ahead.

MEMBER LABOVITZ: I think this is a measure which if publicly reported is immediately understandable to the public. Everybody gets death. But this is not a measure that's able to account for stroke severity. There is no way to grasp at that. It doesn't really account for hospital practices as far as comfort measures, end of life care, transferring patients out.

I'm not satisfied that I really heard enough about how it handles a hospital's choice to accept patients coming in or choice to send them out. That is -- I think Michael Kaplitt's point on that is well taken. It's going to get lost in the model in some -- because it's a relatively small number of patients in a large cohort, but it may be really driving your death statistic. I'm just not sure that it -- I worry that some of our tests for significance here, say an interaction term, are way, way too stringent.

In the end there's a lot of difficult decision-making and a lot of aspects to these models which we don't completely grasp. But in the end what everybody gets is it's about death, and one hospital is going to come out ahead of another hospital. If we don't know really clearly what we're doing with this I think we could take very good hospitals and hurt them.

CO-CHAIR KNOWLTON: Gail?

MEMBER COONEY: Well, as a consumer I just went on the Florida website and pulled up the mortality data that appears to be from this measure because it's using the 3M software. And the one hospital with the higher than expected death rate is our Safety Net hospital and they define higher than expected as more deaths than expected given how sick patients were. Just -- that's what one consumer was able to pull off the website.

CO-CHAIR KNOWLTON: Anybody else? Risha?

MEMBER GIDWANI: I'd just like to hear those C statistics for the models when they were disaggregated.

(Laughter)

MEMBER GIDWANI: I'm just -- were we able to get that?

MR. GEPPERT: I did find them. They were actually much better than I remembered. So the disaggregated, they were 0.88 and 0.87 when they were disaggregated.

MEMBER GIDWANI: Okay, quite good. Thank you.

DR. ROMANO: Just to clarify. So again, those are disaggregated for ischemic and hemorrhagic strokes separately. And those are as good or better than C statistics that have been generated using Get With the Guidelines or clinical registry data. And I think that reflects the fact that we do have proxy measures of stroke severity in the model such as patients who present comatose, patients who present in a persistent vegetative state, patients who present seizing, and so forth.

MEMBER GIDWANI: Yes, those are actually quite good values for discriminating mortality.

CO-CHAIR KNOWLTON: Other comments? Questions? Michael?

MEMBER KAPLITT: Yes, I agree. I mean I think after having raised this point a lot I would concede the point that you know,

while the transfer issue is an important one to me if the numbers, you know, based on stroke type are that relatively well substantiated then presumably it's at least an indirect reflection of you know, more hemorrhagic strokes are going to be transferred from one hospital to another. And so if the numbers are, you know, sort of -- if the numbers based on stroke type are, you know, I mean I'll leave it to the statisticians more to judge that. But you know, I could concede the point that it's at least at some level of, you know, normalization.

CO-CHAIR KNOWLTON: Risha?

MEMBER GIDWANI: I don't think that I handles the issue of transfer out that you brought up, but in terms of transfer in I wonder if the admit risk of mortality would cover that.

MEMBER KAPLITT: Yes, that's what I'm conceding. I mean I think that that's

fine. So I think that, you know, on the down side it protects the hospitals that are taking the sicker patients in transfer. I still think it may artificially inflate the hospitals that are doing less. But you know, I concede most of that point.

CO-CHAIR KNOWLTON: Anything else? Ready for a vote on usability. Let's vote.

MS. THEBERGE: Three high, nine moderate, eight low, two insufficient.

CO-CHAIR KNOWLTON: Okay. It's a close measure. We're moving onto feasibility.

MEMBER RICHMOND: Feasibility. Our group had two high and one low. All data are available in the electronic health record. It uses administrative data.

CO-CHAIR KNOWLTON: Questions? Comments? Seeing none let's vote on it.

MS. THEBERGE: Fourteen high, six moderate, two low.

CO-CHAIR KNOWLTON: Okay.

MEMBER RICHMOND: So endorsement,

our group originally unanimously said no. We had three nos. I can't speak for our group. I'm very satisfied with the information that we got from the developer. It answered a lot of questions that I had and I spent an inordinate amount of time with this measure. So I have converted myself to yes on this.

CO-CHAIR KNOWLTON: Risha?

MEMBER GIDWANI: I'll say the same thing. My "no" was really based off of insufficient information. The developer adequately answered my questions and I feel comfortable so I would now in light of the new information change my response to yes.

CO-CHAIR KNOWLTON: Other comments? Thoughts? Okay, let's vote.

MS. THEBERGE: Fifteen yes, seven no.

CO-CHAIR KNOWLTON: Well done, progressed well through a relatively complex questioning. We move on.

CO-CHAIR TIRSCHWELL: All right,

the next measure. Risha, can you take us through 2026? Okay, sorry. I'm told that we should give the developer a couple of minutes to introduce the measure. Go ahead and start anytime you're ready.

DR. BERNHEIM: Hi, this is Susannah Bernheim. I am a physician and researcher at Yale Center for Outcomes Research and Evaluation, and we work -- louder? Sorry, okay. And we are working under a contract with CMS and we're bringing forward two measures today. I think we're talking about our risk-standardized 30-day mortality measure first. And I'm here with Lein Han from CMS, Jeph Herrin who's one of our statisticians, Judy Lichtman who is a stroke epidemiologist and Elizabeth Drye who will join us shortly.

I'm going to say just a couple of very quick words about the measure itself and then a few words about -- you still can't hear me. I apologize. Okay. Better? A couple of

quick words about the measure itself and then just one minute on risk adjustment for these measures.

So this is a risk-standardized measure. It is a 30-day mortality measure. It evaluates mortality, all-cause mortality following ischemic stroke at the hospital level.

We -- for risk adjustment we are using claims data and we are able to assess patient risk looking both at the inpatient claims and all of the inpatient and outpatient claims for the 12 months prior. So we have historical data on each of the patients.

The model uses a hierarchical modeling that allows us to account for case mix and clustering. And Karen spent nice time this morning talking about that so I'm not going to spend a lot of time on that.

Our measure considers transfers of care as a contiguous hospitalization. We -- okay, sorry. Sorry, okay. Better, okay. For

patients who are transferred between one hospital and the other. The hospital where the patient is admitted is considered accountable for that patient's mortality outcome.

However, in consultation with an amazing set of neurologists who consulted with us on this measure we looked very carefully at patients were seen only outside ED prior to being admitted to the hospital. And based on those evaluations we have added a risk adjustment variable that assesses whether a patient was transferred from an outside ED and that is now risk adjusted for in this measure.

I want to take just one minute to talk about claims data and explain what this measure is meant to do and why the claims are adequate to that task. The measure as you know is designed to profile hospitals. So conceptually we are trying to understand the quality of care through the lens of patient outcomes. And to do that we need to consider

those outcomes in the context of the patients that come into the hospital.

What we are not trying to do is build a prognostic tool for individual patient outcomes. And this is a really important differentiation and I just want to take one minute on it. To predict an individual patient outcome is a different task. We are aiming to adequately assess the risk of the full group of patients that come into a hospital in order to have sufficient confidence that the remaining variation that we see is attributable to quality after we account for uncertainty.

And what we have learned over time is that the administrative data can do that well. What you need to do this well is variables that are consistently collected on all of the patients. And we have the benefit in fact of also having information that's historical on these patients. And which allows us to prevent some gaming.

Whereas a particular variable might be critical for individual patient prognostication, it may fail to be a good risk adjustment variable for hospital profiling. We've learned in this measure and others that the administrative data can produce results that are very close to what is achieved with a model that uses medical record data.

And we've learned that by this measure building the best administrative model we can and then comparing the results at the hospital level with the model that's been built with medical record data. So not only does our administrative claims model achieve a C statistic that's quite comparable to medical record models, but most importantly it's not really in this case about perfect patient prediction, it's about whether you're assessing the hospitals correctly. And we can do that by validating with a medical record model. So we had the advantage of having a national medical record model that we could

compare the results of our model with to be sure that we were determining the same information about a hospital as we would with chart data. So I think that's a really important concept that I just wanted to lay out at the beginning.

As people know, there is wonderful literature coming out of the stroke community indicating the usefulness of the National Institute of Health Stroke Scale for patient prognostication. But sadly we're very far away from having national reliably collected data here. And so our task is to determine whether we can do it well enough with administrative claims.

And we're quite confident that we're bringing forward a model that has a level of patient discrimination that equals many chart models and has a very strong correlation with a medical record model.

MEMBER GIDWANI: Okay, thank you. All right, so to start the panel discussion

I'll give a brief overview. Can everyone hear me? Yes? Okay.

This is a 30-day all-cause mortality rate following an acute ischemic stroke hospitalization. The measure applies to patients who are 65 years and older, and mortality is defined as death from any cause within 30 days of the admission that had the principal diagnosis of acute ischemic stroke.

This measure and the risk adjustment method was based completely off of billing data, ICD-9 codes. It has a number of exclusions from the denominator. Patients who are transferred from another acute care hospital will not be included in this denominator. Patients who have inconsistent or unknown mortality status or other unreliable data, folks who were discharged alive and against medical advice, and patients who were enrolled in the Medicare hospice program at any time in the 12 months prior to being admitted for acute ischemic stroke are

also excluded from this measure.

This is an outcomes measure and the predictor variables and the covariates that are used in the risk adjustment are all patient-level factors. The risk adjustment method is a hierarchical logistical regression model. The hierarchical component of that allows the -- to account for the fact that there are some similarities in patients that are within the same hospital. There's going to be clustering of observations and so the hierarchical component models that aspect.

We had quite a lot of discussion within our work group about this measure. There were four work group members, three of whom voted. In terms of the impact all three work group members rated this as high.

CO-CHAIR TIRSCHWELL: Any comments or questions about impact? Let's go ahead and vote on impact. Go ahead and vote now.

MS. THEBERGE: Twenty-one high, one moderate.

CO-CHAIR TIRSCHWELL: Okay. Why don't we move onto 1c which is evidence as is relevant.

MEMBER GIDWANI: Okay. This is a health outcome measure.

CO-CHAIR TIRSCHWELL: I don't know that we need to go into any further detail. Any comments or questions about that? Let's go ahead and activate the voting. Go ahead and vote.

MS. THEBERGE: Twenty-two yes.

CO-CHAIR TIRSCHWELL: Great. And then onto 1b which is evidence of a gap.

MEMBER GIDWANI: In terms of evidence of a performance gap the developers presented information on deaths by age, gender, race, ethnicity and SES. They show that there is -- that the rate of adverse outcomes and complications associated with stroke increases with advanced age. They note that the overall death rate for stroke is higher amongst black patients compared with

whites. They note that the stroke incidence rate is higher for men compared with women at younger ages but not at older ages.

And in terms of SES they did not see a risk-standardized mortality rate difference across SES quintiles of hospitals. The data that they are showing on disparities by population group are about the outcome of mortality rather than the difference between observed to expected mortalities so I'd like to point that out.

With respect to the work group evaluation, let's see. For performance gap all three members voted to rate this as high.

CO-CHAIR TIRSCHWELL: Any other questions or comments about performance gap? Let's open up the voting then. Go ahead and vote now.

MS. THEBERGE: Twenty high, two moderate.

CO-CHAIR TIRSCHWELL: So then scientific acceptability, reliability first.

MEMBER GIDWANI: With respect to the reliability two work group members rated this as medium, one group member rated this as low.

If I can make a comment here, the developer showed reliability statistics showing the agreement between the risk-standardized mortality ratio for each hospital. The administrative data set was 0.4 which is considered moderate.

CO-CHAIR TIRSCHWELL: Any other comments or questions about reliability? Let's go ahead and open the voting.

MS. THEBERGE: Three high, eighteen moderate, one low.

CO-CHAIR TIRSCHWELL: Okay, next is validity.

MEMBER GIDWANI: There was quite a conversation regarding validity and this is where really the crux of most of the conversations and the questions the developer were posed.

The work group rated -- we had three scores. Two folks rated this as having insufficient evidence. One person rated this as having high evidence of validity.

CO-CHAIR TIRSCHWELL: Okay. I'll start with a question for the developer. You said that specifically you have some patient chart-abstracted data that you used as sort of a gold standard to compare your assessment to. So the hospital ratings which are sort of where the rubber hits the road on this whole thing were highly correlated, sort of the order of rating was correlated between the model and the one based on the theoretical gold standard based on chart review? Could you respond?

DR. BERNHEIM: Right, that's right. We did that validation and the correlation between the chart model output for each hospital and the administrative was 0.8.

CO-CHAIR TIRSCHWELL: So that's the chart model output using the same

specification or you used additional patient-level detail like NIH Stroke Scale score too?

DR. BERNHEIM: Right. So the way we develop the chart model is de novo essentially. We take the variables that are available in the chart model and create a new risk adjustment model using medical record data.

And so then we profile hospitals based on the medical record data-based model and the administrative-based model, and we look at how closely the results of that model for each hospital are correlated. And so we're learning the same information about each hospital, well, to a 0.8 based on one model and the other. I'm confusing you, I can see --

CO-CHAIR TIRSCHWELL: So you're just correlating the predicted mortality between one model and the other?

DR. BERNHEIM: Correlating the risk-standardized mortality rate between the

two models.

CO-CHAIR TIRSCHWELL: How's about comparing the ratings of the hospital? You line up all your hospitals from Connecticut --

DR. BERNHEIM: The same thing. That's what we're doing essentially.

CO-CHAIR TIRSCHWELL: So what's the correlation?

DR. BERNHEIM: 0.8.

CO-CHAIR TIRSCHWELL: That's the r or the r-squared?

DR. BERNHEIM: That's the r -- r-squared.

CO-CHAIR TIRSCHWELL: 0.8 is the r-squared? That would mean that your r was 0.9 which is fantastic.

DR. BERNHEIM: You can see -- it's in our technical report. You can see it listed with the correlation coefficient of 0.8.

CO-CHAIR TIRSCHWELL: Okay.

DR. BERNHEIM: It was the r. I

was correct the first time.

CO-CHAIR TIRSCHWELL: Any other questions that people have about that? I have another question about validity for the developers.

So you know, looking at your list of variables that sort of stay in your model which is extensive, you know, in your introduction you talk about conditions present on admission. So -- and I don't see any real clinical adjusters. So you know, it was previous 12 months plus the index admission. But again, I'm sure that you were careful not to, you know, include anything that might be an indication of poor quality care. And I don't see anything like coma or anything that might be a marker of severity. So can you comment on whether those things were tried and they just didn't stay in the model?

DR. BERNHEIM: I think there's two questions embedded in there so I'm going to take one at a time.

So yes, we're very careful that the risk adjusters that are used from the index admission don't include complications of care. At the time this model was developed POA indicators were not adequate to that task so what we do is we create a list of risk adjustment variables that if they are only present during the index stay may represent a complication and we do not risk-adjust for them unless they are also present historically.

So if a patient has a history of renal failure we would adjust for it if they only appear to have had it during the index admission. We would not use that as a risk adjuster. That's how we handle the complications issue.

I think your second question was where are things like coma. So, the administrative claims do not have a stroke severity scale. There are some indicators. We use a condition grouper that is a CMS

condition grouper, and so some of these individual variables you can't see. A few of them are embedded in there, but again going back to my earlier remarks, what we find is even without those indicators when you aggregate at the hospital level we get an adequate sense of the risk of those patients coming into the hospital.

CO-CHAIR TIRSCHWELL: Okay. And then one final thing about the -- there's a large number of comorbid medical conditions which seem to be paradoxically protective against a prediction of death. And I guess I don't understand how they stand up to the face validity criteria.

DR. BERNHEIM: Yes, this is something that often confuses people and we've spent some time thinking about it. So hypertension is a classic example here that confuses people.

What you need to remember is that we're looking at not just a blood pressure

value as a patient walks in the door with a stroke but historical data. And you'll find even in chart models that if you're looking at this this way the hypertension often is a marker of sort of being less severely sick because it's what's managing to get coded and that's how we interpret that. And we see that a lot. I mean, in the aggregate again these risk adjusters work very well, but some of them because they are historical chart data can seem somewhat paradoxical.

CO-CHAIR TIRSCHWELL: I guess it still doesn't quite seem to meet the face validity criteria. It just suggests that it adds power to your prediction. It doesn't -- I just --

MEMBER J. BAUTISTA: I think what she's saying is that the patient saw a doctor and got diagnosed with hypertension and so is actually being treated. And that's --

CO-CHAIR TIRSCHWELL: I get what she --

MEMBER GIDWANI: I sort of didn't really understand the response to the first question. So you're saying that -- what happens if a patient comes to a hospital and they're a transfer patient and we don't have a history on them because our hospital is in California and those patients live in Arizona.

CO-CHAIR TIRSCHWELL: These are Medicare data. It's national. It's all together.

MEMBER GIDWANI: Okay, all right. Okay, fair enough.

So, I'm also hoping we can bring up Table 9 on page 30 of the Methods Report so that everybody can get a chance to look at the coefficients and really to go off of what David brought up of some of these paradoxically protective conditions.

And I also didn't understand in terms of let's say coma or cerebral edema, mass effect, altered consciousness, those weren't variables that were included in your

model. Were they not originally included in consideration, or were they considered and then dropped out of the model because they weren't statistically significant?

DR. BERNHEIM: So, I think you're bringing up the table there. So essentially every ICD-9 condition code is considered. We use a grouper that collects them into CCs. Sometimes that makes it hard to see individual ICD-9 codes that you're looking for.

We categorically exclude some that aren't relevant to Medicare patients like pregnancy, but otherwise all however many thousand ICD-9 codes are considered as candidate variables. And then you see they each are listed as a CC. So some of these things you're looking for are going to be embedded in other CCs. We have some that have to do with disability such as hemiplegia.

But again, I think the more important piece here is that this is very different than a chart model. We're not using

just a few key variables that are showing up in a chart when they're arriving.

MEMBER GIDWANI: But what this is showing then is that something like cerebral edema is not having a role to play in predicted mortality, but history of infection or major psychiatric disorders is contributing to the risk of mortality.

DR. BERNHEIM: Though I suspect the cerebral edema is embedded in one of these. We could find out, right? I mean again these are grouped variables.

MEMBER GIDWANI: Okay.

DR. BERNHEIM: Each of these CCs represents tens to hundreds of ICD-9 codes.

MEMBER GIDWANI: Okay. And then also in terms of cerebrovascular and cardiovascular this is saying that aneurysm is protective against mortality, that circulatory defects or congenital cardiac defects are protective against mortality. So I too am having a hard time with the face validity of

this.

DR. BERNHEIM: There are certain things that get forced into the model because the clinical experts that we worked with felt that they were important to include in the model. So you're see some of the ones you're pointing to there the confidence interval is crossing 1 and so they don't all actually come out as statistically significant.

I would say this is one of the pieces of our model that committees typically struggle with and it is I think probably where we were 8 years ago. What we have learned in that time is that in aggregate these models do a very good job of assessing the risk of the patients that are coming in. They stand up against chart models in case after case.

And now we are having the benefit of doing more study and learning from some of our measures that have been in play for longer that when you go into the hospitals that do well in these models you see different

characteristics. And we haven't had a chance to do that with stroke yet but we have done in other conditions.

MEMBER GIDWANI: So one thing I'd also like to point out is that correlation between an administrative-based model and chart review may be very high, but if those models are both doing a poor job of predicting they can still have poor predictive ability and high correlation. So, they're just -- would then be doing an equally poor job of predicting.

In this case the ROC statistic, the C statistic or the area under the ROC curve is I believe 0.80 which is reasonable. It's not great but it's certainly reasonable. To put that in perspective a C statistic of 0.5 would mean that the model has no discriminative ability.

DR. BERNHEIM: Can I make one quick comment on the C statistic issue?

CO-CHAIR TIRSCHWELL: Yes, go

ahead.

DR. BERNHEIM: Again, we've talked a little bit about why it's not the only thing that matters in this model. But I also would caution the committee that high, high C statistics can mean that you're really absorbing a lot of the hospital's impact, right?

I mean, we suspect that a patient's outcome is related partly to the risk that they bring into the hospital and partly to the care they give. And that's why it's really important that we do not risk-adjust for those things that may be complications of care. We lose in that case our ability to understand the hospital's impact on a patient. And so a high C statistic does not always mean that the model is better performing. It can easily mean that the model is essentially absorbing that which you're most looking for. So you need to have some caution in this.

The chart models are often in the 0.7 to 0.8 range, the ones that are published. So you know, both our administrative model and our chart model are right in that standard range.

CO-CHAIR TIRSCHWELL: Anything else? Yes, go ahead, Karen.

DR. PACE: I just wanted -- I didn't bring this up when we were going through my initial slides, but just in terms of performance metrics and risk models, to keep in mind that for risk models we're purposely only including patient factors present at the start of care. So, you are going to have different benchmarks on these model performance compared to, say, a total explanatory model where you might be including the care provided and elements of quality of care. So just to kind of keep that in perspective of, you know, we're only including those patient characteristics in risk models.

CO-CHAIR TIRSCHWELL: Any other

comments? Greg, go ahead.

MEMBER KAPINOS: I just wanted to make a comment about like when you were talking about cerebral edema and coma. So, those are abstracted from the billing, right? Not from the coding of the complete notes. And because ischemic strokes are pretty severe, usually taken care by the ER physician, then maybe a neurointensivist, a neurosurgeon, a neurologist, a vascular neurologist. And some systems of billing are limited to four ICD-9 codes that you can bill for. I am familiar with many intensivists actually restricting the number of codes that they use so that actually the other team can also bill for the same patient. And it's not uncommon to have, as I said, three or four physicians billing on the same day for the same ischemic stroke patient.

So very often then in my practice I have not coded a lot of cerebral edemas and comas because they were already with an

ischemic stroke and a respiratory failure. And my system does not allow me to bill for more than four codes.

So I just want to hear back, I mean hear like what's the validity of like abstracting the severity of the patients from ICD-9 codes on the billing system as opposed to just the notes. And even if we use the notes with the DRG and all those fancy models to try to capture the severity of the patient it has -- many clinicians complain that it is still also very imperfect because actually the way you -- whether you dictate your notes or a lot of people are just handwriting or typing does not translate really well. There's sometimes like if you say "pulm edema" instead of "pulmonary edema" that's not going to be charted -- that's not going to be coded for your patient. So there's a lot of -- there's a lot of things that make the system of DRG or billing with the ICD-9 code extremely imperfect.

And from -- I'm junior, so I cannot really -- I want to hear from other, more senior clinicians to confirm that there is actually some good degree of validity to use those billing or DRG system to capture the severity of our patients. Because my understanding is that it's extremely imperfect and so therefore there would be no validity in those models that we're talking about.

CO-CHAIR KNOWLTON: Jolynn, do you want to respond?

MEMBER SUKO: I think we're getting confused. These are based on facility codes. So these are the bill that your hospital submits for the nursing care, all of the other care. And that's typically done in a centralized fashion by coders.

I would agree with you that probably on the physician side when you think about the variation of practice, some physicians are employed, some physicians are in private practice. It's going to be

different. But these are not based upon the codes that you -- they are based upon your documentation but they're not based upon the codes that you as physicians submit on your Part B billing slips.

CO-CHAIR TIRSCHWELL: Although, I mean for the acute care admission I think that's true but there's all this outpatient data from the previous 12 months which are more related to physicians.

MEMBER SUKO: Right.

CO-CHAIR TIRSCHWELL: Gail?

MEMBER COONEY: I just have a question about the exclusion of the Medicare hospice patients and why. Well, my first question started to be why it was only an exclusion on day one, but now I understand that that's because that's all we're looking at. But why is that not a measure of frailty that you would want included in your model?

DR. BERNHEIM: I just want to make sure I understand your question. Why do we

not risk-adjust for hospice as opposed to exclusion?

MEMBER COONEY: To exclusion, yes.

DR. BERNHEIM: It's an interesting question. I think the feeling of the clinical experts was that as opposed to being a frailty marker it was really a marker that in these patients a mortality outcome was not an appropriate measure of quality.

CO-CHAIR TIRSCHWELL: High mortality was almost inevitable probably.

DR. DRYE: Hi, Elizabeth Drye from Yale. Another way to think about it is that this outcome measure is judging hospitals based on, you know, whether their patients live or die. And when we have a patient already enrolled in hospice when they're admitted, then it's clearer that their goal is not necessarily survival. So that's why we don't put them in the measure there. But a different goal instead of risk-adjusting for them.

CO-CHAIR TIRSCHWELL: Risha?

MEMBER GIDWANI: First off my question is -- one of my questions is these estimates, are these log odds that are being presented? These coefficients.

DR. BERNHEIM: The fourth column there, yes.

MEMBER GIDWANI: No, the estimate.

DR. BERNHEIM: Right, the fourth column is the odds ratio. The --

CO-CHAIR TIRSCHWELL: First column is the --

DR. BERNHEIM: The first column is log odds, right. And the standardized estimate is the standardized estimate.

MEMBER GIDWANI: Okay. Again, I'm going to make the suggestion that all data that are presented for coefficients in the future be presented as probabilities. Even odds ratios can be difficult to understand.

More so than that though I have actually a few comments and questions. One,

I'm going to bring up again this issue of where patients are discharged or where they're coming in from. That's not accounted for in these models. If there's a higher risk of 30-day mortality from somebody who's been Life Flighted in that's not going to be taken into account here. If there's a higher risk of mortality for patients who were discharged to a nursing home versus to their home that wouldn't be taken into account here.

I understand the limitations of what you can use from billing and administrative data but my larger concern stems from the fact that CMS has moved to value-based purchasing and that there is a move towards CMS being instead of a fee-for-service provider being a fee-for-value provider and that mortalities and readmissions are a part of their move and that there are financial penalties as well as financial benefits associated with having different levels of mortality and readmissions compared

to the expected level of mortality and readmissions. So I think it's really important that we get these models right given their potentially large implication in the future.

CO-CHAIR TIRSCHWELL: You can respond if you like.

DR. BERNHEIM: I'll just say two quick things. We purposefully don't risk-adjust for where a patient goes again on the principle that that has something to do with the care that's being provided and those decisions reflect that quality. So to the extent that we're not making the right decisions about where to send people that should be reflected in the differences among the hospitals.

In terms of where patients are coming from you're right, we can't do the adjustment for a Life Flight, but again we did with careful consideration and input from our clinician group make sure that we were

adjusting for patients who are coming from an outside ED which will handle some of that issue.

As to the implementation question my understanding, and NQF can speak to this better, is that this is a measure that's been designed for public reporting and this group is here to evaluate its scientific acceptability in that setting. But I would leave that to NQF's guidance.

DR. PACE: I just wanted to confirm, you know, the discharge -- where the patients discharge to would be something that's a factor after the start of care. So risk models should include patient characteristics at the start of care, not things that happen during or at the end of care.

So, and in terms of where we're at now it's about the validity of the measure as it was specified and documented. So, you know, if you have specific questions about

that, you know, as Helen said earlier you know, we have one set of criteria and we expect measures to meet those criteria.

Obviously the Measure Application Partnership does recommend measures that will be used by CMS in a variety of programs including the payment programs. But we do need to focus on your questions about the validity and, you know, obviously that relates to what you're talking about.

CO-CHAIR TIRSCHWELL: Okay. Any other comments on validity before we go to a vote? Okay, let's go ahead and open up the voting for validity.

MS. THEBERGE: Three high, thirteen moderate, five low, one insufficient.

CO-CHAIR TIRSCHWELL: Okay. Moving on next to usability.

MEMBER GIDWANI: With respect to usability the work group was divided. There was one person who rated this as high, one person who rated this as medium, one person

who rated this as insufficient stating the questions about validity need to be settled before answering this question.

CO-CHAIR TIRSCHWELL: So, okay. Sounds like -- and I don't know this for sure, but perhaps after some of the answers that were received the "insufficient" might not be insufficient anymore. Any comments or questions about usability?

Let's go ahead and open the voting then about usability. One response short. Could everybody just hit their button one more time. There we go.

MS. THEBERGE: Four high, eighteen moderate.

CO-CHAIR TIRSCHWELL: Okay. Moving onto feasibility.

MEMBER GIDWANI: With respect to feasibility these are all data based off of the administrative billing record. There was one person who rated this as high and two people who rated this as medium.

One of the comments were that the required data elements, i.e., mortality, do not seem to be routinely gathered nor is there a data collection strategy in place. Another person said the measure is not in operational use but all elements are part of the electronic health record. I'll remind NQF panel members that these are based off of ICD-9 billing data.

CO-CHAIR TIRSCHWELL: Any comments? Questions about feasibility?

DR. BERNHEIM: I can just comment.

CO-CHAIR TIRSCHWELL: Yes, go ahead.

DR. BERNHEIM: Medicare is extremely good at collecting mortality data and that's been validated. So I think the mortality concern for the Medicare population is not a concern.

CO-CHAIR TIRSCHWELL: Okay, thank you. Let's go ahead and open the voting.

MS. THEBERGE: Fourteen high,

eight moderate.

CO-CHAIR TIRSCHWELL: And then finally overall suitability for endorsement. Risha, any final comments?

MEMBER GIDWANI: There were two people who voted no, one person who voted yes.

CO-CHAIR TIRSCHWELL: And again that was before you got the substantial amount of clarification?

MEMBER GIDWANI: That's correct. One person said this is a preliminary conclusion and another person said, "I would like further information and discussion about the presence or absence of stroke severity as part of risk adjustment prior to supporting endorsement."

CO-CHAIR TIRSCHWELL: And I think we've heard about really all of those issues. Any other comments or questions?

Let's go ahead and open the voting then for overall suitability.

MS. THEBERGE: Eighteen yes, four

no.

CO-CHAIR TIRSCHWELL: Okay, thank you. Sure. Okay. Everybody take a deep breath and we'll move onto a very similar measure in some ways, different in others, 2027: Hospital 30-day All-Cause Risk-Standardized Readmission Rate Following Acute Ischemic Stroke Hospitalization. Same group developed it and Risha will again be presenting.

MEMBER GIDWANI: Thank you. This is measure 2027 submitted by CMS. The measure looks at the hospital-level outcome of readmission following an acute ischemic stroke hospitalization for patients aged 65 or older.

The readmission rate is risk-adjusted and is it also all-cause meaning that any readmissions count, even those unrelated to stroke.

The measure does exclude admissions for patients who had an in-hospital death. These patients are of course not

eligible to be readmitted. It also excludes patients who are transferred to another acute care facility. In that case if there were any readmissions it would be attributed to the second hospital that the patient was transferred to.

It also excludes patients who were discharged alive and against medical advice, and excludes patients without at least 30 days post-discharge claims data because they need that amount of information to assess whether the readmission occurred or not.

Again, the level of analysis is at the facility level and this is based off of administrative claims data.

CO-CHAIR TIRSCHWELL: so starting with impact.

MEMBER GIDWANI: Yes, one second please. In terms of the impact there were two persons voting high, one person voting medium.

CO-CHAIR TIRSCHWELL: Any questions or comments on impact? Okay, let's

go ahead and open the voting for impact. Two short.

MS. THEBERGE: We are short two.

CO-CHAIR TIRSCHWELL: Oh, we should only get 21. Oh, we're missing somebody over there as well? Okay, so then we're good.

MS. THEBERGE: Seventeen high, three moderate.

CO-CHAIR TIRSCHWELL: Okay, moving onto 1c. I guess this is an outcome measure but I guess we still need to vote one way or the other for evidence. Any questions or comments about evidence?

Okay, let's go ahead and open the voting. Oh yes, Greg. I can't hear you.

DR. PACE: We do consider it a health outcome because it's really a proxy for deterioration in health status. So, generally we categorize readmission measures as health outcome measures.

MEMBER KAPINOS: But the trigger

to be readmitted can be very low so there could be no deterioration, just self -- I mean, to me no, that's not really deterioration. That's not absolutely the direct measure of morbidity or mortality. Therefore it's not a health outcome.

DR. PACE: That's why I said we consider it a proxy, but we do classify it as a health outcome requiring risk adjustment, et cetera.

CO-CHAIR TIRSCHWELL: Let's restart the voting for evidence. I think that's it if there's still someone missing down there.

MS. THEBERGE: Sixteen -- I'm sorry, nineteen yes, two no.

CO-CHAIR TIRSCHWELL: And then moving onto performance gap.

MEMBER GIDWANI: With respect to the performance gap developers presented information showing that there was a median hospital readmission rate for stroke patients

across the country of 14 percent. They noted there's a large variation in outcomes for readmission with rates ranging from 10 percent to about 19 percent, and those data represent the 25th and 75th percentiles.

They also looked into disparities by population group with population group being defined as race or SES, noting though, however, little work has actually been done on these populations. With respect to race they were not showing racial disparities for African-American patients. With respect to SES they looked at these disparities by looking at the proportion of patients that have dual eligible patients, meaning that they are Medicare and Medicaid, and found that compared to the national average hospitals with higher proportions of dual eligible patients did not have worse 30-day risk-standardized readmission rates.

With respect to the work group panel's -- the work group's evaluation of the

performance gap, two persons voted high, one person voted medium.

CO-CHAIR TIRSCHWELL: Any other comments or questions about performance gap? Let's go ahead and open the voting. Go ahead and vote now.

MS. THEBERGE: Fifteen high, seven moderate.

CO-CHAIR TIRSCHWELL: Okay, reliability.

MEMBER GIDWANI: For reliability the work group members voted two medium, one high.

CO-CHAIR TIRSCHWELL: Any questions or comments about reliability? Let's go ahead and open the voting and go ahead and vote now. One vote short. There we go.

MS. THEBERGE: Ten high, twelve moderate.

CO-CHAIR TIRSCHWELL: Now, validity.

MEMBER GIDWANI: Validity again is where this work group had the most intensive conversation with developers. There were quite a lot of questions about the validity of the measure as specified. One person voted this to have medium validity. Two people voted this to have insufficient validity.

There were a number of questions that were posed by the work group members. Developers did respond to many of those questions. They also did refer us to a report that was convened by CMS. They asked the presidents of all statistical societies within the United States to review their risk adjustment models. I read that report. The presidents did review these models and found them to be appropriate with respect to methodology. And that means that the statistical approach is appropriate.

I do still have some questions about the clinical aspects of the model. I'll allow other people to ask their questions but

-- and I'll save mine until other folks have a chance to speak.

CO-CHAIR TIRSCHWELL: Okay. Michael?

MEMBER KAPLITT: So, with respect to the clinical aspects let's talk about your exclusion criteria for a second. Given the intent of this particular measure, so why -- for example, why are not, you know, people who are readmitted for completely irrelevant reasons excluded? If somebody comes back let's say 3 weeks later with a newly diagnosed cancer. I know that sounds crazy but I'm just trying to use, you know, an obvious example that has nothing to do with their stroke outcome.

And a corollary to that is also a planned readmission. So for example, let's say someone gets a hemicraniectomy because they are swelling and you take off their bone plate. I personally would wait 3 months but some people if they do very well might want to

do it in let's say 3 weeks. We would obviously want to encourage people to let people leave the hospital and then come back rather than encourage them to keep them in the hospital for no good reason just so that their statistics look better. So why were those things not part of the exclusion?

DR. BERNHEIM: I'm going to take them in reverse order because they're easier that way.

The planned readmissions that you point out are excluded. We should have made that clear to the committee. So, we went through with our clinical experts and discussed any likely follow-on procedures that would be scheduled as follow-on care, the largest one being carotid endarterectomy obviously. But we did include cranioplasty in that. That list is on page 11 of our technical report. And so those readmissions are excluded as long as they are not accompanied by a primary discharge diagnosis

that suggests that this was an acute readmission. So if you come back with another stroke then it would be included.

The question about cancer diagnoses, or people's favorite example is car crashes is a good one that we get a lot. And it's -- we feel it's really important to look at an all-cause unplanned readmission for a couple of reasons.

One is that from the patient's perspective this is what affects them. But more importantly except in the very rare case of the car crash it turns out to be really impossible to differentiate what's been related versus unrelated. You know, patients come back septic and it may have been something related to the line that they had in the hospital. It may not have had anything to do with it. And people have done a lot of work with chart measures trying to see if you can parse these things out and you can't.

The important question is is one

hospital likely to have a much higher rate of these kind of random things, and we think that that's really unlikely. We're not in any way suggesting all readmissions are bad or that readmission rates should be zero. We're looking to see whether given the case mix you have, you have a higher than expected rate of readmissions. And in that case we think these other issues pretty much fall out.

MEMBER KAPLITT: I think that's probably true. It's something that I would think should be testable, right? I mean you're right, it should be a minority you would think, although you could make an argument that a level one trauma center which is also a higher level of care for other reasons might be disproportionately affected by it. But I still agree that it's probably an extreme minority but I would think you should be able to generate that data to show that it's a minority, right? That it's not affecting your measure. You should be able to

get data on that point I would think.

DR. BERNHEIM: So, I'm not sure exactly -- so we've worked with colleagues who have done -- to try to do this at a chart level. Are you suggesting with the administrative claims model see how many come in? I'm not totally sure how we would do that exactly.

MEMBER KAPLITT: Well, I mean so for example -- well, I mean, I don't want to sit here and work it out with you because we don't have the time for this. But you know, you could do a prospective study where you look at readmissions, right? For the same patient in the same hospital looking at the diagnoses, and then identify certain diagnoses that might require more further, you know, intensive chart review to be able to get a sense of whether or not this is a serious problem. I mean, I agree that it's probably a minority so I don't want to waste an hour on this, but.

DR. BERNHEIM: No, and there have been some studies like that. And again, they mostly indicate that it's a challenging task to parse.

CO-CHAIR TIRSCHWELL: So I have a question for the developers and it mostly relates to the C statistic of 0.6. And I guess, you know, in my simple perspective if your C statistic is 0.5 then the information that you're giving as the result of your model would essentially be kind of random noise, if it's 0.1 then you're speaking the truth and everything else in between is a gradation.

At 0.6 it seems like there's a lot of noise that must be coming out from the results and given the amount of noise that must be in there the fact that you're grading hospitals based on this, that there's public reporting that might influence patient behavior. You know, I guess to me with a model of 0.6 it's hard for me to justify using that information which, you know, I don't want

to say -- "unreliable" isn't the right word necessarily, but it just seems like how can I really rely on that.

And certainly an end user consumer without any appreciation of that sort of lack of discrimination will only be looking at the end result and they'll take it as gospel. And so I really struggle with how this is really valid for determining much at all.

DR. BERNHEIM: So I'm going to go back to something I said earlier which is remember that we're not attempting to predict a patient's likelihood of readmission. We're trying to understand what's happening at a hospital level. So a 0.1 would mean this was a useless measure because it would say that there's no difference between hospitals. Everything's explained by the patient characteristics when they walk in the door, right? So a 0.1 is not a helpful measure that should not -- I mean, a 1, I'm sorry. I don't mean 0.1, a 1, right?

So, what we understand about readmission is that in fact the patient characteristics don't add a lot to model. Now, the hospital's readmission rate tells you a lot about what's happening to the patients and patient characteristics add a little bit of information there. They do tell you something about --

CO-CHAIR TIRSCHWELL: I just want to interrupt for one second. It seems like you're using this low C statistic, a crutch that that implies that it's more the hospital-related factors when in fact I would submit that your information is much more imperfect in predicting this. And it's not just that the hospital factor is a greater effect. So I don't really think you can use that as a crutch for why your C statistic is low.

DR. BERNHEIM: So I can say a few more things about that. I mean, a number of models have looked at readmission rates. There was a recent review of them. Nobody

finds patient factors are particularly good predictors of readmission. So I mean you may disagree that it's hospital factors, but it does not appear to be patient factors. You can look at it in a lot of different ways, not our models alone although we've now done this a number of times.

The other thing is just to separate the signal-noise reliability thing, we look at reliability in a different way and we do find that we evaluate a hospital time and time again similarly, right? So the signal about the hospital is the same time and time again. We take half the patients randomly and we assess the hospital with half the patients and then we assess the hospital with completely different set of patients and find the same information.

This is the challenge for people with the readmission measures. I mean, the other thing I will say is so one, we don't expect that patient factors are actually

driving readmission rates that much. We think it has much more to do with care, transitions, communication, follow-up and all of those things that we're really trying to spark improvement in.

And we are seeing now more and more studies coming out that in fact hospitals make really important patient-centered improvements and readmission rates drop impressively. And I think it doesn't speak to the C statistic but it does speak to the ability of these systems to really improve the patient experience and allow people to stay home.

Do you want to add something?

DR. DRYE: Yes, I was just going to add talking about it in a slightly bigger picture way. The goal of risk adjustment is really to level the playing field for hospitals, right? Based -- by adjusting for their patient characteristics. And so that's what our model is doing. The C statistic is

just a patient-level statistic. It's the patient-level analysis in the model. And so we know whether we use chart data or we use claims data we can level the playing field, that is we can put all the hospitals on a level playing field, but compared to mortality and other clinical outcomes, readmission rate is -- you're never going to get a good C statistic. What you've accomplished is you've made the measure fair for hospitals.

MEMBER WADDY: So, going back to the comment before last, it does seem like the -- it does seem like patient statistics or patient characteristics can certainly play a role. Is this me?

CO-CHAIR TIRSCHWELL: Could the people on the phone please mute or not step outside of the airplane anymore?

(Laughter)

MEMBER WADDY: Such things as, you know, if a patient is for some reason non-compliant with their medications and they come

back in with pulmonary edema or they decide not to take their antithrombotic then they could come back in with a stroke. And so I mean those things are very complicated and they certainly can also be tied into hospital characteristics as well, but they aren't mutually -- they're neither mutually exclusive nor completely encompassed by evaluating at the facility level.

CO-CHAIR TIRSCHWELL: Response from the developers?

DR. DRYE: Sure. That's a really good point and I just try to separate a little bit further the goals of risk adjustment which is we're evaluating what we're doing to level the playing field, and other patient characteristics. We're assessing -- for the model, the risk adjustment model we're just looking at patient clinical characteristics and demographic, when they arrive in the hospital. And so we're not addressing patient behaviors.

And in the context of readmission measures we're -- there's a lively discussion about patient behaviors and certainly patient behaviors influence risk of -- I don't, you know, I can't give you how much they influence it but I agree with you that they would influence risk of readmission. More compliant patients are less likely to come back.

And that's one of the myriad of factors that we understand hospitals can influence. They don't have full control over it but they can influence it by medication reconciliation, clear discharge instructions providing better support post discharge. So it's a factor hospitals can influence that we don't want to -- we wouldn't want to adjust for it anyway. Yet we, you know, agree they don't have full control over it.

CO-CHAIR TIRSCHWELL: I guess, you know, some of the arguments you make about the hospital and the patient factors. And I don't know the literature or whether it exists on

this. So if you did go ahead and designed a perfect model that included those factors I guess you're suggesting that those C statistics would be vastly higher. Is there any evidence of that?

DR. BERNHEIM: You mean if you designed a perfect model that accounted for nursing care, communication, collaboration, appropriate discharge care? I mean, I don't think anybody can design that model.

CO-CHAIR TIRSCHWELL: But even partway there?

DR. BERNHEIM: I mean, I think if you add complications into the model you would probably learn something but you erase part of the signal. I mean, that we have seen.

CO-CHAIR TIRSCHWELL: Done that.

DR. BERNHEIM: We haven't done that. That has been done in other settings where the complications of care.

But a lot of the things that we believe are really influential are not easy to

measure individually which is why in this case an outcomes measure is so important for quality improvement because there aren't, you know, the process measures that have tried to get at this have had a very hard time discriminating against -- between truly good care and not. And so I think that's the gap that this measure helps to fill.

CO-CHAIR TIRSCHWELL: But it seems there's an implicit assumption that the hospitals can have a big effect and you're judging the quality of the hospital care on their, you know, that they have the ability to really affect this readmission rate. And I guess I don't know that that's -- I don't know that that's true. That's a bit of a leap of faith. And you know, whether the C statistic is the right way to try to determine that or not I don't know. But I'm just --

DR. BERNHEIM: Right, I mean these are -- sorry, go ahead. No, no, it's two different questions I think. But I would re-

frame the implicit leap slightly. We do know that hospitals can influence this because we are starting to see evidence. And they, you know, those studies are percolating out in a lot of places. And we do know that hospitals currently have not focused on these key components which lead to increasing the risk of readmission that have to do with patient education and reconciliation and really communication across providers and coordination.

CO-CHAIR TIRSCHWELL: Don't say patient education because --

DR. BERNHEIM: Oh, sorry.

CO-CHAIR TIRSCHWELL: We disavowed that measure yesterday.

DR. BERNHEIM: Forget I said that word.

CO-CHAIR TIRSCHWELL: All right, let's go to some more comments. Mary?

MEMBER VAN DE KAMP: I just had a question. The re-hospitalization rate is an

aggregate now for hospitals and yet we're looking at it for diagnosis for stroke. Is the intent then to look at a re-hospitalization rate specific to the kinds of diagnosis or discharge?

For instance, you'd have a different re-hospitalization rate for an ischemic stroke than you would have for cardiac? I'm confused maybe in the overall intent long-term. It's specific to diagnosis, is that what you're saying?

Looking at it -- like I spent a lot of time in long-term care facilities and looking at re-hospitalization. And I'm looking to see if what the patient is currently -- diagnosis is is impacting a different rate. So you're going to have a higher re-hospitalization rate for a stroke than you would for pneumonia.

DR. BERNHEIM: So I want to make sure I understood your question. The measure looks at patients whose initial admission was

for an ischemic stroke and evaluates whether they have any unplanned readmissions regardless of the cause. But I'm not sure that I answered your question.

MEMBER VAN DE KAMP: So as you take that out to the discharge location and now you have, you know, a skilled nursing arena that has a number of different diagnoses that, you know, hospitals are looking at for re-hospitalization rate. Is the rate benchmark going to be different per diagnosis for that re-hospitalization rate?

DR. BERNHEIM: The way that this measure is designed the benchmark will be against other patients who had an ischemic stroke hospitalization.

CO-CHAIR TIRSCHWELL: This one is specific to the diagnosis at the time of hospitalization? I think later in the competing discussion there's a more general readmission one. All right, never mind.

MEMBER VAN DE KAMP: I think in

the environment that we're in right now we hear re-hospitalization rates kind of generically thrown around as what, you know, what's your re-hospitalization rate. And they don't pull it apart per diagnosis.

CO-CHAIR TIRSCHWELL: Well, just the measure that's before us is specific to ischemic stroke.

MEMBER VAN DE KAMP: Right, so I guess I'm asking -- I guess I'm probably in usability now that I'm talking that way.

CO-CHAIR TIRSCHWELL: Okay. Okay, thank you.

MEMBER VAN DE KAMP: Sorry.

CO-CHAIR TIRSCHWELL: Dan?

MEMBER LABOVITZ: We've had some access to some Medicare data at our hospital and what we've found is that readmission very much depends on things that the hospital does have control over, but not in the way you think. If we pick the right subacute rehabilitation facility, the readmission rate

is much lower because that rehab facility has a doctor on staff, has a system that works well, and may not be the most expensive one but it's better at taking care of its patients. And if the hospital picks the right place your readmission rate is lower.

Hospital has no incentive to pick the right place right now. It has the incentive to pick the place that will take the patient fastest. And I think what we're really talking about here is that stroke is not something that's just an episode of care. It's not the hospital admission. It's a process that unfolds over weeks and months. But probably the place that has the most impact on how that unfolds is the hospital where the patient starts.

My hospital gets to pick which rehab facility the patient goes to and I think we're just beginning to look at that. But what we're discovering is yes, our quality is pretty bad there because we're running after

a dollar in this direction. We're not motivated to go after it in another direction.

I think this is a good start, but I do have some concerns. There are things that --

CO-CHAIR TIRSCHWELL: Excuse me, sorry to interrupt. Can the people on the phone just mute their lines, please? Or Operator Amy, can you mute those lines? Thank you. Dan, sorry.

MEMBER LABOVITZ: I think there are some things that I maybe haven't studied this enough or didn't spot that I think really do have an impact on risk of readmission that need to be accounted for lest we hurt hospitals that are doing a good job or introduce bias.

I think there are community-based factors that influence a patient's risk of readmission. And I'm wondering if that's been looked at, if that's been evaluated here. Even zip code of origin might have an impact,

and I think insurance status makes a difference.

I know that the patients I discharge home with really abysmal support because they're undocumented and I can't get them a thing. But I can't keep them in the hospital forever because the vice president of finance will call me the next day. Is that accounted for here? And that's not the hospital's fault, but it is the hospital reality.

CO-CHAIR TIRSCHWELL: Thank you. Ramon?

MEMBER R. BAUTISTA: One thing I like about your measure, it actually has at least a sincere attempt to try to level the playing field across different providers and hospital systems. And being it is admission there's partly a leap of faith maybe here in accepting this measure. And maybe a few years from now we can find out if it really works or not.

A relatively minor question though. Because it is Medicare data we're looking at you can actually trace readmissions to another hospital that takes place, right? And ding the first hospital in that regard, right? Okay, thank you.

CO-CHAIR TIRSCHWELL: Therese?

MEMBER RICHMOND: I was on the work group and I was one of the "insufficient" people. I'm feeling more comfortable with this measure.

While I don't think that it is a perfect measure by any stretch of the imagination, I do think that looking at hospital readmissions there's a growing science in models of care that could reduce readmission in patient populations, congestive heart failure, vulnerable older adults, et cetera. So I think from I guess both importance but also a validity perspective this is probably an important measure.

CO-CHAIR TIRSCHWELL: Michael?

MEMBER KAPLITT: Yes. So you know, to Dan's point I would just like, you know, an answer. It seems to me like your risk adjustment is pretty similar to the last measure except risk adjustment for death is very different than risk adjustment for readmission. Because there seems to be no accounting for the post-discharge risks which there are which is very different than death which, you know, we understand. That's more reflective of your baseline risk. So he mentioned a whole series of them. So, was there a discussion about post-discharge risks? And you know, and if so then why are they not accounted for?

DR. BERNHEIM: So I thought Dan had some really good comments that got at the nuance of this issue, right? So on the one hand, you know, it's not our fault, it's the rehabilitation center. On the other hand, we are the ones who are in the position -- "we" being a hospital -- to evaluate the post-

discharge setting and make sure that when we have provided excellent care to patients we are not then sending them to a place that is going to unravel that.

And so it is not a simple situation and the way the health system is designed right now is very segmented, siloed. And hospitals can't control everything that happens. But if you think about a nidus for this measure and who is in the best position to take accountability I think there will be growing efforts to have community agencies also taking some responsibility. I think you will see that and may already. We feel like the hospital has an enormous ability to affect this and we are seeing that.

MEMBER KAPLITT: But with all due respect, I mean I agree with you that, you know, you don't want to oversimplify it but the whole exercise here is somewhat oversimplified, right? Because we're going to give people a single number that says that

this hospital may be better at this than this hospital. But then we're saying that, you know, well, you know, we understand that hospitals have issues and they can control some of this.

And I think that you're way overstating what the hospital can control because it varies by state. There are state laws that influence things, right? Patients, families have a right to make a choice as to where they want to go even if we disagree with them. You know, there are the realities as he says about different patients having different financial situations that influence their post-discharge care. We can give two patients the exact same level of care and yet what happens to them afterwards really has very little to do and very little control by the hospital.

So what we're doing is saying okay, in a world where the hospital is put in a situation where they have limitations on

what they can do, we're going to ignore factors that could influence this and make the assumption that the hospital has more control than they may have.

And I think that's the concern that's being expressed here. Not that we disagree that this is ultimately not an important measure, we've already voted on that. But the question is how we're taking into account the realities of the world we live in now rather than the way we'd like it to be.

DR. BERNHEIM: Right, so I don't mean to oversimplify or be unsympathetic. There is no question that the causal pathway to readmissions is incredibly complex. There's not an easy way to try to tease apart those factors in the post-discharge environment that a hospital can or cannot influence, and it is clear that there are many things that hospitals can do that will reduce risk.

And hospitals again are not being expected to go to zero. The question is given the case mix you have how are you doing relative to other hospitals. And I think in that way we're leveling the playing field and doing the best we can in an environment where this is a really important measure.

CO-CHAIR TIRSCHWELL: Jack, do you have a comment?

MEMBER SCARIANO: In private practice we have what's called -- it's called hospital wars. That if you look in a city you always see billboards up. It says this hospital is actually number one in heart, or this hospital is also number one in heart. And they all use different criteria.

Well, I've seen in our city that oftentimes to have an overall better heart rating is that heart problems often get dumped into stroke. If you come in and you have a heart failure you may get confused and oftentimes the cardiologist would say well, it

was an actual stroke that actually came in. And as he got bad and as he or CMS came in and actually did an audit and the hospital is now closed because they were doing this.

So I think that the overall way you can tease this out is to have like a quality committee in the hospital look at all the readmissions and see, you know, what's the actual cause. You know, is it actually heart failure? Is it actually a kidney failure? Is it dehydration? Because all these things at times are actually being logged in as they're having stroke.

CO-CHAIR TIRSCHWELL: Okay, thanks. Helen?

DR. BURSTIN: So this is obviously not the first readmission measure NQF has looked at and I suspect it's probably not the last. So I just want to at least give some insights into where this has gone before.

So, some of you may know we recently evaluated the all-cause hospital

readmission measure and as part of that discussion the board did exactly this discussion. We had this discussion for several hours and ultimately the board put out a guidance statement that I think might be helpful just to put this in context when they ultimately endorsed that measure.

And the point was multiple factors affect readmission measures including the complexity of the medical condition and associated therapies, effectiveness of inpatient treatment and care transitions, patient understanding and adherence to treatment plans, patient health literacy and language barriers, and the availability and quality of post-acute and community-based services particularly for patients with low income. Readmission measurements should reinforce national efforts to focus all stakeholders' attention and collaboration on this important issue.

So I think there is a recognition

readmissions are multifactorial, there are many factors that go into play. And I think the recognition was measuring it at a hospital level will probably enhance more of the community collaboration that I think really you just talked to, Dan, in terms of understanding what are available in terms of community resources and others.

So I just wanted to put that on the table. This isn't a new issue but it certainly is something we've spent a lot of time talking about over the last few months.

CO-CHAIR TIRSCHWELL: Okay. Risha, final words?

MEMBER GIDWANI: Before I begin speaking can I ask that we bring up Table 10 on page 36 of the methodology report?

So, I share a lot of David's concerns with the poor C statistic and that to me is actually quite concerning. And developers, your response to us is saying that the rationale for the poor discriminative

ability is hospital-level characteristics, but I just don't see the data for this. I'm not sure whether this is really a poor C statistic because the model isn't accounting for enough patient-level factors or because it's correctly excluding hospital-level factors. And I'd like to see some evidence in support of your statement that it's because of the lack of hospital-level factors.

DR. HERRIN: So first, what are we looking at on the table that concerns you?

MEMBER GIDWANI: Well, these are all of the patient-level factors that are in your model and so I want the clinicians in the room to be able to see this. I'm not a clinician but maybe this is -- if this is considered a comprehensive list by our clinicians I'm happy to go with that, but if there are some patient-level factors that are not included here then that would point towards the need to include these in the model rather than state that the poor discriminative

ability is completely due to no hospital-level characteristics being included.

CO-CHAIR TIRSCHWELL: Along those lines I guess I'd ask has anybody done a similar readmission model where you had, like for the mortality one you said you compared to ones with, you know, stroke severity nicely coded and characterized. And did that lead to a better ability to predict readmission?

DR. BERNHEIM: So we did the same chart validation for this measure and the medical record model actually had a slightly worse C statistic. And they were correlated at 0.99.

CO-CHAIR TIRSCHWELL: Did the stroke severity -- was that a significant predictor of readmission at all? Because I think that's one of the main things that's missing from administrative data.

DR. BERNHEIM: Right, that's the concern that people raise. When you look in the literature which is not yet deep on what

predicts readmissions and stroke, it turns out that stroke severity is only variably showing up as an important predictor which is surprising.

CO-CHAIR TIRSCHWELL: Risha?

MEMBER GIDWANI: I just wanted the developers to respond to the original question.

CO-CHAIR TIRSCHWELL: Go ahead.

DR. HERRIN: So we're talking about the C statistic. And I understand that the 0.60 looks low but if you think about the fact that we're measuring hospitals and the first thing you might do to measure a hospital is just calculate the raw rate. Take the number of readmissions and divide it by the number of patients, you get a percentage. You can use that as a model also to predict what happens to each patient. And if you do that the C statistic is -- at the patient-level is something like 0.52. I mean it's not much better than chance. I think we'd all agree

that at the hospital level you actually have a pretty good first, you know, first order estimate of the what the hospital rate is.

All we're doing is taking that rate and adjusting it. It may not look like we have a very good prediction at the patient level but I think that what we end up with a hospital rate is, you know, is an improvement on just the raw rate. That's what you want.

And the fact that we reach 0.6, the comparison is not -- we're not trying to again predict what happens to individual patients. We're trying to measure what is happening at the hospital level.

CO-CHAIR TIRSCHWELL: Risha?

MEMBER GIDWANI: I think my concern stems from the fact that given that this is a poor C statistic, if it is entirely -- the distance between 0.6 and 1.0 is entirely due to hospital-level factors, okay, that would be information for the hospitals if you know, any data were actually being

collected and presented to them on this but I think that's a usability issue rather than validity. But it seems to me that you have that opportunity, that if you have the medical record you could actually include the hospital-level factors of the transfer process and some aspect of communication and other variables that you consider to be hospital-level and see whether using your medical record model your C statistic was greatly improved.

And if that was the case then all of your other variables, your patient-level variables were the same between your medical record model and your administrative model. And the only difference was the inclusion of hospital-level information in the medical record model. Then I think you could make that conclusion, that the poor C statistic is due to the lack of hospital-level information.

But given that the opportunity existed to do that we don't see any data here.

I'm just concerned. I think that that statement that's being made, maybe it's true but the evidence isn't there to back it up.

CO-CHAIR TIRSCHWELL: Response?

DR. BERNHEIM: I'll just point again to the fact that there's been a number of models, not by our group but by other groups looking at this and consistently this is the finding when you try to look at the hospital-level readmission patient factors adjust. Patient factors that you would want to adjust for, patient factors that are present on admission do not seem to be particularly important in evaluating. So I mean I just, I know that makes the committee uncomfortable but it seems to be true across the board.

I think what you're proposing we could try to look at is to assess what variables at the hospital level we could collect that we think might be important. You could also look at things outside of the

hospital potentially. It's a challenging project. I mean again, I think we all believe that what is predicting readmissions is really a complex web of missed opportunities to coordinate care.

And so to adequately try to put all of those into the model is not a simple job and I think it's why people haven't done it. But we can go back to our group and think a little bit about whether there's ways that we can sort of prove that hospital factors being added would increase the C statistic if that would be useful.

CO-CHAIR TIRSCHWELL: Yes, and you know my --

DR. DRYE: I just want to clarify then the goal there is not to change the measure because as we talked about we -- patient factors do matter. We don't, like Jeph was pointing out, the C statistic is not 0.5, it's 0.6. It's important to adjust for patient factors on admission the way we do.

We would not be fair to hospitals if we didn't do that. So, all our models, our chart-based model, our claims models, other people's readmission models, this is as good as they get with the C statistic.

So to clarify what -- for readmission. So what you're asking is, you know, can we do an investigation that gives you more confidence that hospital factors that we can get our hands on influence the outcome of readmission. And it's kind of an ancillary study saying is readmission really --

MEMBER GIDWANI: I'm sorry, I don't think it is ancillary because the response to our concern over the low C statistic was that it's because hospital-level factors are appropriately not included. And so I think that's an important thing to test before stating that that's the reason.

CO-CHAIR TIRSCHWELL: It's sort of the foundation that the potential improvement is -- I mean, the idea is that you're going to

improve care here by affecting the hospital-level factors which are the explanation. And I guess, you know, we've held other measures up to show the evidence that what we're trying to influence here, hospital-level care, is clearly shown to affect the outcome of interest.

DR. DRYE: I want to -- I think my colleague Harlan Krumholz is on the phone and wants to say something. But by "ancillary" I don't mean that it's irrelevant, I just mean it's not about changing the measure, it's about thinking about readmission as a measure concept. And I think the kinds of things you would want to get at, we've been talking about, like medication reconciliation, coordination, rapid response, complications in the hospital because those affect readmission, you know, safety, and then all the transitional and post-acute care. They're not sort of easy, quick things we could grab for hospitals and throw into the model to look at

that, but I want to -- let me let Harlan follow on.

DR. KRUMHOLZ: Thanks, Elizabeth. And I appreciate -- I'm sorry.

CO-CHAIR TIRSCHWELL: Go ahead.

DR. KRUMHOLZ: Okay, thanks. This is Harlan Krumholz and I'm a member of the Yale team. And I appreciate the opportunity to speak to the group. Can you hear me clearly?

CO-CHAIR TIRSCHWELL: Yes.

DR. KRUMHOLZ: Great. This issue of course as Helen has said has come up repetitively about the C statistic and it's one that we have thought deeply about. It defies easy empirical analysis by putting in hospital interventions because of the heterogeneity of the way in which these interventions are applied.

See how teaching, for example, discharge instructions doesn't turn out to be a very good measure and doesn't turn out to

indicate at all any better outcomes. Yet all of us believe that really good teaching likely has a role to play in helping improve patient outcomes and that's because when you study care you see immense variability in the way in which that process is applied. So it becomes very difficult to take hospital characteristics and -- them at covariates in this model and try to explain some of the variation because they're all complex.

Now with regard to this issue of the low C statistic, there -- I think -- I just want to review what are some of the major points here. One is that remember we are purposely tying the risk adjustment to admission because things that happen in the hospital, adverse events that happen in the hospital we would not want to adjust for and give a hospital credit for a sicker group of patients because of -- complications that may occur in the hospital many of which may be preventable in a lower risk environment. So,

it is one of the things that makes it almost by nature going to be able to predict readmission is the distance in time from the time zero with this discharge. So that's one thing.

The second thing is that no matter what data source we have used around readmission we continually find that patient characteristics are far from the dominant influence on who gets readmitted. It has led to this appreciation of thinking about people leave at a certain risk strata. Each environment is associated with risk.

And when we look deeply at our own institution we find an embarrassingly high number of opportunities here to improve. That is, we find that we are often sending people home -- I won't say often, but we found we were sending people home with two beta blockers. We had forgotten to give the antibiotic for the patient who was admitted with pneumonia. We have given people a liter

of fluid the day before they go home after heart failure. We have failed to give them a path towards appointments after they leave. We've done a lot of things that actually we think increase their risk of readmission and in fact for the three publicly reported measures we've historically been higher than expected. And we have instituted a lot of new approaches and our readmission rate is dropping.

I spoke to 600 Premier hospitals just 2 weeks ago in Nashville. And as I made my way around the room I'm hearing about people recognizing that they've got these same deficiencies. And they know they don't own the entire 30 days, and they know that there are many things that are beyond their control. But they're also seizing the things that are within their control and they're recognizing not that they can eliminate readmissions, but they can lower the risk of patients, make it safer for them to go home, be more secure in

that the systems that are being implemented are going to smooth that path for those patients. And that's what makes us feel confident about these measures.

It's true that it's difficult to predict. It's true that there are many things that are beyond a hospital's control. But when you look deeply at the hospitals we do not do well at this. And this is now shining the light on it and hospitals that are higher than expected commonly have more problems. And we're seeing people being able to make movement by focusing on that. And that's why we remain strong in our belief that this is an important measure.

CO-CHAIR TIRSCHWELL: Thank you very much. Salina?

MEMBER WADDY: So, the first question I'd like to ask is in response to David's question a few minutes ago regarding severity. Since this is really, as the caller mentioned, the time zero is date of discharge

is the severity the --

CO-CHAIR TIRSCHWELL: Day of admission I think.

MEMBER WADDY: No, no, according to this it says date of discharge.

DR. BERNHEIM: We only assess patient factors up till the time of admission but the 30-day time window starts at discharge.

CO-CHAIR TIRSCHWELL: Oh okay, sorry.

MEMBER WADDY: Right, that's why I was wondering if the severity that you're -- that you've looked at, is that severity at admission which it sounds like. It seems like it would be more appropriate to have severity at time of discharge.

CO-CHAIR TIRSCHWELL: Well, they don't really have severity at either time.

DR. BERNHEIM: But the concept that you're trying to get at, again, we are trying to understand a multitude of factors

that are going to affect the likelihood that a patient's going to get readmitted. Some of those happen during the hospitalization and those are the ones that are more under a hospital's control.

So if a patient is more severely ill at their time of discharge, let's say we failed to do aspiration precautions and they've ended up sicker we would not want to risk-adjust that away. So the risk adjustment's time, if you will the risk adjustment time zero starts at admission but we want to assess a standard period for readmission so that needs to start at discharge so that we don't have variable length of potential readmissions.

CO-CHAIR TIRSCHWELL: Dan?

MEMBER LABOVITZ: I'm just offering up some thoughts in response to Risha's points which are I think is we're getting back to the C statistic of a lousy 0.6. And I think the developers are

rightfully admonished for suggesting that the remaining distance to 1.0 is just hospital-based factors. I think we just don't know.

But I also would suggest that our capacity to capture these things in models as the -- I don't remember his name, but the developer who just spoke fairly eloquently said you can capture -- you can record all kinds of data but it doesn't necessarily have meaning, and it doesn't necessarily have anything to do with quality or what you're really delivering.

And I think one of the advantages of this look is it's two hard points. You've got a hospital and you've got a real rubber-meets-the-road readmission rate. And I think that there are very, very significant -- there's a real capacity for the hospital to influence that rate, not totally, not even close to totally. The patient will get hit by a bus, bound to happen, and there are going to be other factors in the community that

influence it. But this is an area where there is tremendous opportunity to improve and I don't really care how you do it. Maybe it's throwing an education pack at the patient, maybe it's writing "You shouldn't smoke" on every discharge summary, or maybe it's establishing better connection between the nursing home and the physician who discharged the patient and making sure that happens. And yes, you get an administrator who can answer the phone and get you to the right doctor, this sort of thing. Hospitals will find a way to do it if we shine a light on it.

CO-CHAIR TIRSCHWELL: Yes, so Dan, if I can just reiterate that maybe it's not the particulars of the model that are the key thing here, it's that we're talking about it at all and that that's leading engagement in hospitals and maybe even beyond their borders to try to reduce these rates. Jolynn?

MEMBER SUKO: Well, a very similar point. But as you think about conceptually an

outcome measure it's to drive the discovery of the interventions that may influence it. And so when you look at this, this is driving discovery even more so than mortality of interventions that we believe will influence it. And so even though the C statistic isn't perfect like no model is, when you look at 2b.5 Meaningful Differences I think we're finding that with this measure.

CO-CHAIR TIRSCHWELL: Salina?

MEMBER WADDY: So just a follow-up on Dan's statements. Certainly there are things that the hospitals can do, some things that are out of their control but there are things that they can do better. Unfortunately there's a paucity of tools that we actually know that work.

And so one thing that's currently going on within NIH is to better develop tools that kind of bridge the gap from the time a patient is in the hospital, the use of behavior change interventions that are not

only behavior change for the patient but also behavior changes for the hospital as well as the primary care provider and the utilization of things such as community health workers to try to solidify the lessons that were supposed to be learned in the hospital.

And so eventually you'll be able to -- people will be able to use these tools and whether or not a hospital system actually adopts one tool or the use of no tools is going to separate out the quality of care between those types of systems hopefully.

CO-CHAIR TIRSCHWELL: Okay, Bill and then maybe last words.

MEMBER BARSAN: Just one quick question about the variables. Did you all look at any other mental health variables besides dementia?

CO-CHAIR TIRSCHWELL: They're probably bundled into those giant bundles I'm guessing.

DR. BERNHEIM: Yes, I'm trying to

remember which ones came into this model. Yes, so again, the mental health variables are bundled into a couple of different grouped ICD-9 codes. And in this case what you see up there is what was found to be consistently statistically significant for the model.

CO-CHAIR TIRSCHWELL: Risha?

MEMBER GIDWANI: I just want to clarify my concern here is not only with the C statistic. I think the developers' explanation of why a C statistic can be low would be valid if they provided evidence to suggest that the entirety of the difference or the majority of the difference between 0.6 and 1.0 is due to hospital-level factors.

And given the fact that they did have an opportunity to study this difficult though it might be, and I acknowledge that it is, but also the repercussions of these measures are quite large and I think that warrants a very thoughtful and careful look. And given the fact that they did have medical

record data and somewhat, well, really inability to actually study this and operationalize this with good resources, but the data weren't presented. I don't think the anecdotal evidence that were presented by the gentlemen on the phone are sufficient for what a National Quality Forum endorsement would require.

DR. BERNHEIM: Can I just respond briefly to what's available in the medical records? Which again, as we're talking about we don't think that the factors are things like is this a teaching hospital or not, right? I mean again, this discussion has been I think a very thoughtful one about the complex web of things that are probably contributing.

And so in order to examine this we would need -- when we have chart data what we have is data that's been abstracted from charts that looks at patient factors. We don't have data that was abstracted that looks

at the quality of the discharge instructions or many of the things that people have been referring to here that might be important.

And again, we can, you know, we can think with you about whether there really are variables that would help to answer this question, but it has not been taken lightly. It's just a pretty tough task.

CO-CHAIR TIRSCHWELL: Any final comments before we move to vote, Risha? Last comment?

MEMBER GIDWANI: I would just say then that points to insufficient evidence as to the difference between a value of 0.6 and 1.0 as opposed to us just concluding that it should be due to hospital-level factors. I don't think there's the data then.

CO-CHAIR TIRSCHWELL: Okay.

DR. DRYE: Can I just add one -- just one? I think it's relevant to this issue of whether there are hospital-level factors that affect the outcome is that there is a

published peer-reviewed literature showing effective interventions by hospitals in lowering readmissions. So we know hospitals can affect the outcome of readmission.

CO-CHAIR TIRSCHWELL: Okay. And final comment from Karen?

DR. PACE: Right. I just want to make a few points about our criteria and some of the points that have been brought up.

One, first of all about our preference for outcome measures and the acknowledgment that the reason we don't ask for the developers to provide all of the detail of the body of evidence like we do for process measures is because there are multiple processes and care interventions that affect outcomes.

And so just as was already mentioned that an outcome measure is never going to tell you exactly what to do. It is something that tells you that you need to dig into your data and see for your particular

setting and patients what are those interventions. And it is through measuring outcomes that we actually push the envelope to try to find those things. So, that is kind of the essence of our board's direction of trying to get at outcome measures.

The other thing as was already mentioned is that there is a growing body of evidence that does exist showing the impact of interventions on readmission rates.

And thirdly is that even though the, you know, the C statistic, you can't say that it's all going to be hospital factors, that certainly is part of the explanation. But one of the things to look at in any of these statistics because again our criteria have not set kind of a hard threshold that a C statistic has to be a particular number or a reliability statistic has to be a particular number, but what is it in relationship to norms for that particular outcome or that particular measure or that particular

reliability statistic.

So I just wanted to kind of bring you back to some of the criteria and the discussions and reasons for the approach that NQF has in their criteria.

CO-CHAIR TIRSCHWELL: Okay. Unless there are other burning questions I'd suggest we move to activate the voting for validity.

MS. THEBERGE: Twelve moderate, four low, six insufficient.

CO-CHAIR TIRSCHWELL: Okay. So then next is usability.

MEMBER GIDWANI: With respect to usability the work group had a value of one high, one medium, one insufficient. The rationale for the insufficient was more discussion of how to interpret a predicted to expected value is needed. The developers did provide feedback on this. And I was the person that noted that so I would change my vote.

One question I do have though is that if these are hospital-level factors that would need to be intervened upon in order to reduce the readmission rate I don't believe these are actually being captured. So I suppose, is it then correct that if a hospital had a poor -- higher than expected readmission rate they would need to delve into their own records and do their own analyses to decide how to improve that?

DR. BERNHEIM: Yes, I think there are some -- again, there are increasing evidence about interventions that are useful. But there is an expectation that the outcomes measures really spark a fair amount of work for the hospitals to understand where they can intervene. And again, probably not just within their own walls.

CO-CHAIR TIRSCHWELL: And I guess I would only comment that I remain concerned about the interpretability of a ranking based on something that's not very predictive of the

outcome.

DR. BERNHEIM: Can I say one quick?

CO-CHAIR TIRSCHWELL: Please.

DR. BERNHEIM: It's -- Elizabeth was going to go back to the -- again, we're not trying to predict patient-level. But also I would comment that the way these measures have traditionally been used really is not as a ranking, right? I think it's important to know that the way that they have tended to be used in public reporting is simply to identify outliers. So, to identify the hospitals that are doing significantly worse than would be expected given their case mix, but not to sort of say hospital A is one point better than hospital B is one point better than hospital C. That's not the way they've traditionally been used.

CO-CHAIR TIRSCHWELL: So you're just ranking the worst ones as worst.

(Laughter)

DR. BERNHEIM: Sorry, just identifying outliers. Just an identification of outliers.

CO-CHAIR TIRSCHWELL: Seems like a ranking. But anyway, that's fine. Risha?

MEMBER GIDWANI: I should have asked this question earlier but if a hospital has a value of let's say 1.2 but their confidence interval goes from -- includes 1.0 then would they be considered average?

DR. BERNHEIM: Yes.

MEMBER GIDWANI: Thank you.

CO-CHAIR TIRSCHWELL: Any other comments or questions? Mary.

MEMBER VAN DE KAMP: I'm going to try my question again then. And maybe I think you helped, Gail. The all-cause re-hospitalization metric that was used, how does this differ from that? Are you looking to then say that re-hospitalization rates may be different as you look at different diagnoses? So if I'm taking a lot of stroke patients I

need to be better at that and that would show rather than lumping it into an all-cause bucket that may not really differentiate specialty?

DR. BERNHEIM: Right, exactly. I think -- I didn't realize earlier that you were referring to the hospital-wide measure.

So, we think that they both have a purpose, that the hospital-wide measure may be an important way of looking at a hospital as a whole and that it is likely that there are going to be a number of things that cross specialties that are important and that have a quality signal at the hospital as a whole. But your neurologist group is going to struggle to use that measure to improve care for their patients. And so there's a real need for quality improvement to be able to look at these things at a condition-specific level. So that would be the use of this measure.

CO-CHAIR TIRSCHWELL: Okay.

Risha, your thing is still up there. Do you have anymore comments? Okay. Jocelyn?

MEMBER J. BAUTISTA: One quick question. Can you clarify, would patients admitted under observation status, would they be excluded?

DR. BERNHEIM: Right, the measure is designed to capture patients who are admitted for the index stay as well as for the readmission.

CO-CHAIR TIRSCHWELL: So sounds like they would be excluded.

DR. BERNHEIM: Right, but maybe that got lost. Yes, you are correct, they would be excluded.

CO-CHAIR TIRSCHWELL: Okay. I say we open the voting for usability.

MS. THEBERGE: Seven high, eleven moderate, four low.

CO-CHAIR TIRSCHWELL: And then feasibility. Any comments? Risha, do you want to say anything?

MEMBER GIDWANI: I'll just summarize the scores which were two high, one medium. And I'll remind everyone these are administrative data.

CO-CHAIR TIRSCHWELL: Any comments? Let's open the voting, feasibility.

MS. THEBERGE: Eleven high, ten moderate, one low.

CO-CHAIR TIRSCHWELL: And then overall suitability for endorsement.

MEMBER GIDWANI: For overall suitability the work group voted one yes, two no. One work group member noted a preliminary conclusion that this was preliminary based on more details on the modeling process and rationale. Another work group member wanted further discussion on the inclusion or absence of stroke severity in the risk adjustment and the implications of this.

CO-CHAIR TIRSCHWELL: Any other comments or questions? Let's go ahead and open the voting up.

MS. THEBERGE: Thirteen yes, nine no.

CO-CHAIR TIRSCHWELL: Okay, I think we're going to take a 10-minute break. Let's try to be back by 11:20, we'll get started.

(Whereupon, the foregoing matter went off the record at 11:08 a.m. and resumed at 11:21 a.m.)

CO-CHAIR KNOWLTON: We are moving onto the speech and language measures. And we're going to ask for the developer to make a comment first and then Karen has a comment from NQF. And then we're going to move out of order because Dr. Sheth has a flight he has to catch so we'll be moving 0446 up to the first one. So you can adjust your SharePoints, your documents if you want to. So let's go to the developer first.

DR. MULLEN: Good morning, everyone. My name is Rob Mullen and I'm joined by my colleague Dr. Frymark. We are

with the American Speech-Language-Hearing Association or ASLHA representing the measure development team.

These measures were developed by a team of clinicians and researchers at ASLHA 15 or 16 years ago and have been in use for the past 14 years primarily through an ASLHA-sponsored nationwide data clinician system called the National Outcomes Measurement System. So we have been collecting data using these measures for the past 14 years.

We currently have about 300,000 episodes of speech-language pathology treatment in our data set based on these measures. They are currently used within the context of a National Outcomes Measurement System by approximately 3,000 clinicians and approximately 500 facilities across the United States as well as a smattering of other countries.

With NQF endorsement of these measures I believe it was 4 years ago the

measures that went into the public domain saw certainly additional use beyond the previously restricted use for the National Outcomes Measurement System. So for the past 4 or so years there has been additional use by other people for other purposes.

I think a couple of important things to note is that these measures were developed 15 or 16 years ago not with public reporting in mind. Obviously they've been submitted to NQF for endorsement because we do think they have the potential to be used in public reporting but the initial impetus for developing these measures was to have locally available data for clinicians and administrators to be able to assess and document the functional gains made or not made by patients at the local level to stimulate thinking about quality improvement. And that's primarily how they've been used.

The eight measures here represent the eight areas of speech-language pathology

treatment most commonly used with stroke patients. In practice the eight measures are not used together and we certainly have never seen a patient for whom all eight of these measures were scored. Typically what we see is that the patient will use -- a clinician will use one to maybe two or three or even four of these measures on a single patient, but they are meant to be separate depending on which of these areas of speech-language pathology relate to that patient's treatment plan.

So these measures consist of basically a pre-score and a post-score. At the beginning and at the end of the speech-language pathology treatment episode the patients are scored on these disorder-specific seven-point ordinal scale. So it's important to note that these are ordinal rather than interval scales. And the primary measure of progress we use is the extent to which patients made or failed to make any measurable

progress on these scales from admission to discharge from SLP treatments.

MS. JOHNSON: Thank you. Just to give you a little bit more background -- thank you, Rob, for that intro to your measures. We wanted to give you from the NQF perspective just a little bit more background on the work that we've done between these developers and us to try to get these ready for you guys to look at. So, if you'll bear with me I'm going to just give you that background now.

First of all, the first time around when they submitted their measures we had a lot of questions just like with some of the other developers. And these developers were great in really responding to our questions. So some of our questions included questions about the impact of the measures and also a lot of very detailed questions about their specs. We weren't quite clear about their definition of progress, their time measurement, their exclusions and their risk

adjustment methodology. So we did ask them to provide that and for the most part they were able to do that.

We also asked for additional detail about reliability and validity testing methods and results. So both of those things.

And again, they did respond at length with a lot of things. They told us about impact and they really brought in information from the literature for that. In terms of evidence they, as you know now, these are outcome measures so they were not required to give evidence in terms of quantity, quality and consistency, but they did I think show some rationale supporting their treatment hours I think is how they did it, treatment hours to outcome.

They really precisely specified their measures and in terms of data element reliability that's not something that they had to do because they did show data element and validity. So again, that's an NQF guidance

there that if you show data element validity we don't require data element reliability.

And then finally with their validity testing they have done validity testing at the patient level for the scale that they use, the seven-level scales, and they've also provided some measure score testing, some results from that.

All of that said we still have a few unresolved questions and we just wanted to put these out for you guys to be thinking about as you do the discussion. And I think you probably would have even without this slide, but opportunity for improvement. What is the distribution of the performance scores for the measures as specified? And they specified these measures for both clinicians and facilities.

And what they gave they did -- depending on the measure they maybe have as many -- as few a six or even as many as 24 strata. So they did give you differences in

patient-level scores for those strata as appropriate, but we also would like to see those for clinicians and facilities because that's how they're specifying that they would use these measures.

For reliability if you have as many as 24 risk categories the question there is do you have enough numbers to have, you know, good comparisons.

The risk adjustment strategy, the questions that maybe are still there are the analysis that support the risk categories that they have specified as well as a demonstration that the risk adjustment is adequate.

And then finally probably the least important question but also an interesting one is how these measures as specified compare to what is currently being reported in PQRS. So with that I'm going to stop and hand it back over to our chairs.

CO-CHAIR KNOWLTON: Reminder -- Jane, do you have a comment?

MEMBER SULLIVAN: Just a point of clarification. Karen, were the questions that were asked prior to or after the work group call?

MS. JOHNSON: Some of the questions we asked right before the work group call and Rob and Toby had those answers for us by the work group call and we -- I believe we sent those out to the full committee. And then other ones came after the call.

CO-CHAIR KNOWLTON: Okay. A reminder that we are now considering 0446, Functional Communication Measure: Reading. And Raj is going to present for the work group.

MEMBER SHETH: Thank you. I think the -- looking at the numbers from the impact the group felt that the data that had been provided, the rationale that about 16.5 percent with stroke actually have speech and language services and 25, a quarter of that group failed to make any improvement in

progress. And they also, the other rationale for this is that there's disparities between race and gender as an issue to be dealt with.

The way in which the numerator was scored was really an increase of one or more levels in the reading score. The denominator had exclusions if there was only one visit. And obviously there was no way to measure whether the score went up or down, stayed the same. So the group as a whole felt on the impact factor that this had a high impact and one felt that this was a low impact.

CO-CHAIR KNOWLTON: Questions or comments? David? I thought you were raising your hand.

MEMBER SULLIVAN: I have a point of clarification on the denominator exclusion and I noticed this actually last night. For each of these measures the exclusion says "Patients who are not candidates for memory treatment." And I believe that's inaccurate. I believe that it should be for each of the

areas of care. So this one should be not eligible for reading treatment, is that correct?

DR. MULLEN: That is correct.

MEMBER SULLIVAN: Okay.

DR. MULLEN: I apologize.

CO-CHAIR KNOWLTON: Other questions? Jocelyn?

MEMBER J. BAUTISTA: So, the evidence of high impact is basically that there are 15,000 patients who receive these services, is that right? Is there additional information?

MS. JOHNSON: This is Karen. That is one of the things that they did add to their submission. So if you were looking at the old submission you wouldn't see the stuff from the literature. It should be in there. No? Okay.

CO-CHAIR KNOWLTON: Do we have an open question on that or is it resolved? It's not there?

MS. JOHNSON: Let me pull up that one and check and make sure we've given you the right one.

CO-CHAIR KNOWLTON: While we're doing that, Michael?

MEMBER KAPLITT: Well, I mean mine is basically the same and it's an overarching question that I think is going to be the same thing with each of these because it looks to me from just the few that I've skimmed through that the impact section is pretty much the same from one measure to the next showing the same 15,000 patients about what a big problem it is and then isolating what percent have this particular thing but no real statement of impact as to how each of these specific measures are supposed to impact care. Maybe that data is not in what we're looking at right now.

CO-CHAIR KNOWLTON: A.M.?

MEMBER BARRETT: So I'll comment and perhaps the NQF staff can add. Since this

an outcome measure although we would -- the work group noted that we were concerned about the fact that not many patients have been included in the database that was assessed, that this may be because of the opportunity to further expand the measure rather than because of a limited impact.

CO-CHAIR KNOWLTON: Karen?

DR. PACE: Yes. So, I know that this kind of maybe in some respects looks like splitting hairs, but in terms of impact opportunity for improvement and evidence, what we're kind of looking -- and these are outcome measures. But what are the numbers of -- I mean, first of all you could look at the numbers of people with stroke who have this particular deficit. And so I guess your question is are they giving specific information for the deficit. It doesn't necessarily have to be in their database. This information could come from national studies or -- and you all would be more aware

of the numbers that exist in terms of patients with stroke who have this particular deficit, whether -- go ahead.

MEMBER KAPLITT: So what I'm getting at, for example, is a simple thing. So it says that the numerator I think is an improvement of one point or something on this scale, is that right? So where's the evidence that one point is meaningful and will make an impact and matters? And is that one point equal -- is the scale perfectly linear? You know, I mean that's what I mean by impact.

DR. PACE: Right. So I think we'll get into that in the specifics of the measure. That's about the validity of the measure as being an indicator of quality. So your question is is a one step up, is that really going to be appropriate --

MEMBER KAPLITT: Yes, and maybe it's just a difference of opinion. Like when you say -- to me impact is I want to know what they're defining as being, you know, a change

is going to impact.

DR. PACE: Right. And I'm just telling you in terms of NQF criteria what we're getting at is impact is the potential numbers of people who could be influenced by this particular measure. So we look at impact and then opportunity for improvement. So even though there would be a lot of people affected, you know, if performance is already extremely high then again there's not going to be much improvement. So you're right, you know, with the --

MEMBER KAPLITT: Is there data on that point here?

DR. PACE: I think you're --

MEMBER KAPLITT: I mean there's 2,494 patients, right? How do we --

CO-CHAIR TIRSCHWELL: They say in the updated thing, they say there's a million aphasic individuals in the United States and that that's mostly due to stroke, and 30 percent of stroke has aphasia. Those are the

high-impact numbers I think.

CO-CHAIR KNOWLTON: Dan?

MEMBER LABOVITZ: I guess the problem I have in terms of assessing impact is that I don't see any evidence that relates to impact here. I see that we've got a problem, we have a lot of aphasic patients. I see that 25 percent of aphasic patients don't make an improvement and 75 percent do.

But where's the impact? What are we influencing? What are we changing here? What is -- speech-language pathologists are some of my best friends. I love them.

(Laughter)

MEMBER LABOVITZ: I ask for their help all the time. But I want to see the impact. What are we achieving?

CO-CHAIR KNOWLTON: Let's not crosstalk. Let's do this in an orderly way. Mary?

MEMBER VAN DE KAMP: Dan, I'm your best friend and as a speech and language

pathologist I think what this gets at is beginning to measure the effectiveness of the kinds of treatment procedures we provide to patients. So, until you know that there's improvement made or not improvement made in a certain disability or area of focus you can't look back to say what treatment was provided that caused that patient to do better or what comorbidities caused that patient.

So, until you can start to measure what we would all agree upon would be certain levels of performance you can't go back to look to see what actual procedures were done that got a better outcome than another. So two speech pathologists doing whatever we think is right, we can't really judge or look back to say what was actually the best practice in that treatment. So that's -- the impact is the quality of the speech and language services that are provided. And then as an industry or a company you can start to measure.

It's just like, you know, you look at the care tool that CMS is looking at. It's looking at a tool to measure outcomes. And right now we don't have a standardized measurement for rehabilitation to discipline-specific outcomes to measure is too much speech pathology the right -- I mean, is this too much? Is that too little? Was that right for that patient? Because we haven't standardized as an industry, and this begins to standardize that process. So I don't know if that answers the impact. I think there's twofold for that.

CO-CHAIR KNOWLTON: Ramon?

MEMBER R. BAUTISTA: So it's the measure's intent to have all patients with strokes undergo speech therapy consult and undergo the FCM? Is that the intent of this measure here?

DR. MULLEN: It is not.

MEMBER R. BAUTISTA: So how would we know who undergoes FCM?

DR. MULLEN: The person who should be scored on the FCM is the person in this case for 0446 would be the person who has a stroke who is treated by speech-language pathology typically for a reading disorder. So that is not the same as saying that we think that all stroke patients should be treated for reading. That is not the intent.

CO-CHAIR KNOWLTON: Jocelyn, do you want to add something to this? I see you reaching.

MEMBER J. BAUTISTA: So the way I interpret high-impact in terms of what we need to evaluate is what numbers of patients does this measure impact, right? And is that a large number? That's basically what we're being asked to evaluate here, right? So for this measure, this measure will affect roughly 15,000 patients a year, those patients who receive pathology services. Am I?

DR. PACE: So let me try another way. You're looking at this in relationship

to a specific measure. Forget about the specifics of this measure for a moment and the question is this is the one with functional communication measure, reading. So the question is is this a, you know, does it affect a large number of patients? Is it subject to quality issues? Is there high resource use associated with it? Or is it a, you know, high patient and societal consequences to this issue? So you'll get at whether the particular measure is an appropriate way to address this.

This is strictly a question of whether this is an area we should have a performance measure at all because of those kinds of things. If there's a lot of people, there's really severe consequences or high resource use, et cetera. You'll get to the specifics of the measure in terms of whether that's the way to go in terms of this area. So this is, you know, at a higher level in terms of is this really even an area that

merits us taking a look at and having performance measures.

CO-CHAIR KNOWLTON: Do you want to reply to that, Jocelyn?

MEMBER J. BAUTISTA: So operationally then are we asking ourselves is stroke a high-impact area, or are we asking ourselves are the numbers of patients treated --

DR. PACE: I don't think you can focus it on the number treated because part of the problem may be they're not getting treated. Yes, right.

CO-CHAIR KNOWLTON: Jane?

MEMBER SULLIVAN: I think -- I understood that we were to look at these measures individually and that this measure is about reading deficits. And it's a subset of all these people who have stroke who have communication deficits who have reading. So it's 16.5 percent based on the data set presented that had -- that were treated for

reading dysfunction. So it's a smaller group than stroke or than stroke that has communication problems.

DR. PACE: What I was just saying, it's people that have this particular deficit because as you were saying not all stroke patients will have this deficit. What I'm saying is I don't know this data but they're presenting data from their data set on people that they know who have been treated for this. Perhaps there are more stroke patients who actually should be treated for this. I don't know in your field whether that's the case or not, or whether those who are treated is truly, you know, just the number who have this deficit. So it is about strokes and in this case the reading deficit that we're talking about.

CO-CHAIR KNOWLTON: Just speaking for myself I'll take that on its face but I would like to acknowledge some agreement with Michael's point that I could do a correlation

that says that 75 percent of all stroke patients have brown eyes. You know, it doesn't -- so you could say well that impact is great because it's half of all stroke patients, but there's no relevant intervention, there's no impact there whatsoever because it's not tied to any type of particular outcome.

So it does get a bit confusing if you separate the impact because otherwise I guess you're just doing it on numbers, does it affect a lot of people, and we could have all kinds of things that affect a lot of people that are not relevant to a measure that you would want to put.

DR. PACE: Exactly, right.

CO-CHAIR KNOWLTON: And so I understand your point.

DR. PACE: So I think the overarching thing is that this is all in the context of quality of care for the stroke patients and particularly those who have

reading deficit. But you're right, I mean, and it's an area that we have had other discussions about trying to clarify or collapse these.

CO-CHAIR KNOWLTON: Michael, go ahead. I'll come back to you, I'm sorry.

MEMBER KAPLITT: Because I mean I think we're all kind of saying the same thing it's just I guess the question is like in 1a where is the data that says specifically that a reading deficit is a big problem in stroke? I mean, maybe if that's the simplest way to put it, right? Because -- and the reason I say that is because all of this stuff about aphasia appears in many of the other measures. And so if we're going to just take that as the impact then why are we measuring 12 different things here today? You know, why don't we just have one global measure?

So that's I guess what we're struggling with trying to look at because if it turns out that it's just 2,000 patients or

something then obviously that's not. And if we take it on faith that maybe there's more, you know, well, what's the data?

DR. PACE: Right, no, and that's a fair question.

MEMBER KAPLITT: And I think that's what we'd like to know. So maybe the developer or somebody can give us more information.

DR. MULLEN: One way to put these numbers into context is that the 15,000 episodes of care from last year were those reported to our National Outcomes Measurement System. And our best estimate is that approximately 10 percent of eligible speech-language pathologists who are eligible to participate in this system do. And so then I think we could -- we could generally say that the total number of episodes of care of stroke patients receiving speech-language pathology services is somewhere north of 150,000.

We just have 10 percent for those.

And assuming our data are representative then about 16 and a half percent of those 150,000-plus patients were treated for reading disorders.

CO-CHAIR KNOWLTON: Okay.

DR. MULLEN: So that would put it more in the neighborhood of twenty-five or thirty thousand.

CO-CHAIR KNOWLTON: Helen?

DR. BURSTIN: I just want to make one point that it's also not the absolute numbers. And if you look up there it's also severity. So if we limited everything to just the numbers of people you would oftentimes leave out things that are actually quite serious but maybe don't affect a lot of people.

So, we did, you know, a fair amount of work a couple of years ago on pediatric heart surgery. Again, not huge numbers, but pretty significant impact for those who do. So I just want to at least put

that in context for you.

CO-CHAIR KNOWLTON: A.M.?

MEMBER BARRETT: Let me comment that the work group struggled with this issue that's being discussed of, you know, the feeling of good faith with the developer that what was being presented could fully help us to be responsible to the larger potential scope of NQF endorsement, right? Beyond the NOMS database.

And the guidance we received on the work group call was that for the area of impact we may be able to use our own expert judgment to some extent, and please correct me if I'm incorrect. However, with the other areas like reliability and in particular validity we can drill down much further as the group feels appropriate.

CO-CHAIR KNOWLTON: Other comments on impact? Karen?

DR. PACE: So, did you get an answer to your question about the reading?

MEMBER KAPLITT: Yes, I guess. I mean, what I would really like to see rather than, you know, the percent of patients that were treated is some data from studies that say what the scope of the problem is, this specific problem.

You know, are there studies that say that you know, 10 percent of all stroke patients let's say have specific reading problems where this outcome measure is actually going to make a big impact, you know? That's really not here. That would be nice, and that would be nice for all the other things. I don't get the sense we're going to get that today but that's sort of what I'm driving at I think.

CO-CHAIR KNOWLTON: Ramon?

MEMBER R. BAUTISTA: Is there data -- for the speech people in the group, is there data that shows that an improvement of one point or one level in the FCM can happen without any rehab? In other words, can this

be a natural course of getting better after a stroke? I mean, I don't know the answer to that. I ask the speech therapists here. Do we actually need an intervention for this or would this happen as a matter of natural course?

CO-CHAIR KNOWLTON: Mary, do you want to take a shot?

MEMBER VAN DE KAMP: You're asking the million dollar question. I think that that is a challenge in any sort of rehabilitation to determine if you didn't intervene what would the result be. But to take the chance of not intervention, you know, I think CMS asked that question in payment. You know, if you just left a person to rehabilitate or improve how much is just going to naturally happen with this.

I think that's one of the reasons we can look at outcomes. That's one of the reasons by having an outcome we can start to drill back and look at are those the -- what

are the reasons that it doesn't improve and can we compare. But if what we struggled with in the rehab industry is any sort of benchmark that we would all standardize across each other's provision of services to begin to look at what happened. So I think if we had an outcome to say 80 percent of the patients who had reading issues were treated and improved this much we would have a measurement to decide how they improved that much. Right now without that we can't answer some of those questions that will talk to what Karen said is resource utilization. Because that's one of the things that's looked at. So a long-winded answer to your question. It's a very difficult one.

CO-CHAIR KNOWLTON: Going back -- continue.

MEMBER R. BAUTISTA: It would sound like a placebo-controlled trial would be a more reasonable thing to do rather than having a national measure to require everybody

to do this with no end in sight. I mean, just my opinion.

DR. MULLEN: This is Rob. There does seem to be some indication from our data that there is certainly a possibility that some patients will make a level of progress in the absence -- we don't have data on patients who receive no services, but we certainly do have data on patients who receive very little service, you know, less than an hour in some cases and some of them do make progress.

What the data from the National Outcomes Measurement System shows is that the likelihood of making progress is very strongly related to how much treatment they receive. So there will be some. I think it's probably safe to assume that there will be some who would make progress in the absence of any treatment.

CO-CHAIR KNOWLTON: Therese?

DR. MULLEN: The treatment certainly increases based on the -- increases

the likelihood of making that progress.

CO-CHAIR KNOWLTON: Go ahead, Therese.

MEMBER RICHMOND: Two issues, and this is all in this section. One is I agree, they did not -- you don't really see evidence that reading -- I would have been convinced if we saw this number of people have reading, this is the impact on life. People then are functionally much more impaired in terms of the ability to, you know, carry out normal life activities. So I didn't see that.

And the second thing, and this may be jumping ahead, is I'm not convinced and I don't want to -- I'm not a speech-language pathologist. However, I feel like I'm looking at a 2 by 2 table here that's missing half the table in that we're shown that hours of intervention that we -- the percent goes up, you have an increased percentage of people who improve. However, you know, time is a factor here that's not really controlled for.

So we're only seeing people who are treated, who have an intervention and they progress, but since those interventions happen over time we really don't know whether the intervention is linked to that outcome or the person would have improved just by virtue of time. And I don't see any evidence here that shows linking that structure-process outcome.

CO-CHAIR KNOWLTON: Other? Jane?

MEMBER SULLIVAN: As a rehab therapist I share Mary's sense of, you know, this is a first step in trying to standardize what we do and standardize the way we look at what we do and get some answers.

I guess one of the things that is troubling to me is the percent of clinicians that would be eligible to report on this measure. And as I understand it, only 10 percent of eligible speech-language pathologists have done the training to do this -- these measures. So you know, it's further a small -- in terms of impact it's a smaller

percentage of clinicians and therefore a smaller percentage of patients that it affects. And I know you want to think about driving practice in a good way but in terms of impact that gets smaller and smaller.

DR. MULLEN: Well, I would suggest it's sort of a catch-22 situation in that with NQF -- with continued NQF endorsement that would be important and stimulating increased participation.

CO-CHAIR KNOWLTON: Karen?

DR. PACE: Right. So I think these are all important questions and I guess I think some of these apply to other criteria and so maybe you want to talk about impact. But the question about the relationship to treatment is what we would talk about under evidence.

And you know, for outcome measures we don't ask them to submit all the bodies of evidence but to provide a reasonable rationale that there are interventions or treatments or

services that do impact that, and that's certainly up for your discussion of whether you think, you know, there really is any impact.

But you know, perhaps -- and then certainly how many people are using the measure, you know, under usability that would be great. And I think it is, you know, it's not a requirement for NQF endorsement that people are already using it though this is coming back for endorsement maintenance. So it's certainly a fair question to ask when we get to usability in terms of, you know, why isn't it being used more and what are the plans to really get it into public reporting. So these are all important questions but you may want to kind of move through.

CO-CHAIR KNOWLTON: On impact. Michael?

MEMBER KAPLITT: I mean, the last statement from the developer concerned me because I -- you know, we're not an NIH study

section here and we're not a, you know, a foundation. We're here to have a different purpose is my understanding which is not to figure out the potential of this to do things or whatever, but is there enough evidence to say that people should be measured by this standard now.

And that's where I think the impact question is coming in here, that is there -- have we been provided with enough data to say that this specific measure, that there's enough evidence to say that we should now endorse this or maintain the endorsement, that this is where all of you guys should be measured by. It's not a matter of whether this is important or whether there's the potential or some people could benefit, you know. And that's my concern here is are we, you know, do we have that.

DR. PACE: Right. So your question is very specific about -- and which is brought up about the numbers of people, the

consequence of the reading deficit, et cetera, and that's exactly what you should be focused on right now.

CO-CHAIR KNOWLTON: Therese, is your hand still up? Okay. Any other comments? We're voting on impact. Open the voting, please.

MS. THEBERGE: Four high, eight moderate, four low, five insufficient evidence.

CO-CHAIR KNOWLTON: Okay. I understand that Mary's stepping in for Raj. He had to leave to get his flight. And we're onto evidence. Is there sufficient evidence, importance of the measure evidence, yes or no. Up to you, Mary.

MEMBER VAN DE KAMP: I think as we go back to what Karen was saying the evidence is not as significant a requirement within the outcome process. And so as we talked about is there evidence of this being a risk within this measure we all agreed, those of us who

voted.

CO-CHAIR KNOWLTON: Questions? Comments on evidence?

CO-CHAIR TIRSCHWELL: I just wonder, I'm recalling back to the, you know, assess for rehab measures that we approved the other day. They quoted thousands of studies supposedly showing that rehab had benefit and at least some of them must have included assessment of some of these speech and language pathology services. So, I am not a master of that literature but it would seem to me that there must be some evidence that interventions along these lines which we haven't talked about yet exist.

MEMBER VAN DE KAMP: Rob, do you have something? I don't have -- I have -- the details on mine was what was presented to the work group, but I don't have -- do you have --

MEMBER BARRETT: Well, I'm just going to comment again that we were directed that for the number one criteria we don't have

to be dependent on just information that is presented by the developer. And so indeed I would confirm that the fact that good practice standards exist requiring reading and other speech-language pathology treatments in most high-quality settings would see -- of evidence. Although that wasn't made in the application.

CO-CHAIR KNOWLTON: Anything else on evidence? Karen, is your hand up?

DR. PACE: So, would you put up 1c on the -- so you can see what they presented? So this is the area with the health outcome. Is there a relationship to the, you know, structures, processes, services. And that certainly is something that you can discuss.

And I think they used as a proxy the relationship between treatment service. We only look at section 1c.1. Okay. So go down to the next page is where they provided that information. And that's I think what some people were questioning earlier but this

is the place to bring that question up.

CO-CHAIR KNOWLTON: Therese?

MEMBER RICHMOND: -- my earlier point is I don't think that this is convincing evidence of linking the process of care to the outcome measurement. I think we're seeing only people who were treated with hours of treatment, but there's -- I would like to have seen at least evidence from the literature linking interventions of speech pathology with improved outcomes. So I don't believe that evidence was shown or I'm not seeing it.

CO-CHAIR KNOWLTON: Other comments? Okay --

MEMBER BARRETT: I would just say the work group agreed.

CO-CHAIR KNOWLTON: On the issue of evidence let's vote. Oh, Mary had a point. I'm sorry, I didn't see it.

MEMBER VAN DE KAMP: I think the question that we have is that we didn't have to demonstrate the evidence within this. We

could use our, you know, our research for our disciplines and per the group on the panel that was mixed, not just speech-language pathologists. But I think is that -- are we voting on -- and I guess I'm still confused because I know Karen was great because we tried to really get into this on our work group. We struggled with this one a little bit being that it was an outcome measure and not a process measure and where that definition was. So to Karen.

DR. PACE: So, what we're asking and you know, what we like to see here is for the developer to identify the relationship between at least one service intervention, health care structure that impacts this outcome. So, sometimes we'll see, for example on the readmission measure some discussion that transition practices, discharge status, coordination of care, getting the patient to their right next provider are things that impact readmission and there is, you know, we

don't ask them to go through the same, you know, description of the body of evidence as you saw yesterday for process measures.

So, this developer is noting that speech pathology treatment is related to making progress. You know, that's still open for you to decide whether that is sufficient rationale for the measure. I'm just saying that we didn't ask the developer to submit, you know, a summary of the body of evidence like we require for the process measures.

CO-CHAIR KNOWLTON: Jane.

MEMBER SULLIVAN: I think one of the large points of discussion on the work group call is the fact that this is a maintenance measure and there's been some time since the measure was first endorsed. The work group was looking for some data that would show that using this measure has had some impact, that you know, more than what was perhaps submitted the first time. And we had hoped and thought that the developer was going

to provide that for us. And I'm not seeing that.

CO-CHAIR KNOWLTON: Anything else? Okay, now we can vote on evidence.

MS. THEBERGE: Five yes, sixteen no.

CO-CHAIR KNOWLTON: Okay, so this measure does not get approved for -- re-approved I guess.

We now will go into usual order and will be up to 0442. And David, you're going to do this?

CO-CHAIR TIRSCHWELL: Okay, so Jane, is this one that you presented?

MEMBER SULLIVAN: Yes, this one's mine.

CO-CHAIR TIRSCHWELL: And I guess to some degree we need to reflect on what just happened.

MEMBER SULLIVAN: I think there's going to be similarities throughout this measure. Just one point of clarification. In

terms of the scoring on the document that you received there were three people on the work group call but the numbers are four. I was credited for two votes. I don't know if that's because I'm from Chicago.

(Laughter)

MEMBER SULLIVAN: There were only three people. I'm not sure what's most helpful. I think that the conversation that we had on the last measure is going to be very much like this one. I think the difference is as I look numerator statements, very similar denominator statements, similar with respect in this regard to writing. People who are using an augmentative alternate communication system are excluded from this measure.

If we go down to impact the -- we have the same kind of data. In this case the developer talked about 10 percent of the subset of people who were being seen for speech-language services were receiving services for a writing disorder. And that was

the extent of the impact data that we had to evaluate.

CO-CHAIR TIRSCHWELL: Any further comments on impact? Let's go ahead and open the voting for impact.

MS. THEBERGE: Four high, seven moderate, eight low, two insufficient evidence.

CO-CHAIR TIRSCHWELL: So we proceed. 1c is evidence. Jane?

MEMBER SULLIVAN: I think the findings here that were presented by the developer are consistent with what we talked about in the last measure. There was some information that time of intervention, hours of care does affect outcome but that was pretty much the extent of it.

CO-CHAIR TIRSCHWELL: Right. And -- yes. Any further comments or questions or points of differentiating this measure from the previous one?

DR. MULLEN: As the developer if I

could just say that I think that there is no cause to differentiate the evidence -- here or any of the ones for the remaining measures from the one that was just addressed. So I guess, I don't know if it would be some sort of violation of NQF protocols but if the previous measure will not be moving forward because of the evidence criterion, it's not going to be any different for this or the remaining measures. So I don't know if there's some way to speed up the process so no one's time is wasted with individual deliberations of the remaining measures. Because the evidence sections are approached in the same way across these measures.

CO-CHAIR TIRSCHWELL: Thank you very much for that comment. Karen, did you have something to say?

DR. PACE: I think before -- I just want to bring out this question to the committee and also to the developer. You chose to present it this way in hours of

treatment, but those of you in the field, are there -- is there evidence of specific speech-language treatments that do impact this outcome? So, you know.

MEMBER VAN DE KAMP: I guess I'm going to go back. I'm sounding a bit like a broken record, but we do use it for that purpose. But it's just like the FIM scoring which was used in the stroke study. They measured what the FIM change was and then they went back to find out what was the procedures within that change where they had a greater change. Were there different procedures used to better determine the best practice of that care? And so that's how we used that within our company. But that's a different -- that's not publicly --

DR. PACE: Right, I understand that's how you used any outcome measure --

MEMBER VAN DE KAMP: Right.

DR. PACE: -- in terms of determining how to improve. But often

generally there is some evidence to start with of even giving speech-language pathology treatment. Is there some studies that indicate that certain types of interventions actually impact patients?

CO-CHAIR TIRSCHWELL: A.M.?

MEMBER BARRETT: Mary, you can fill in, but certainly one of the professional societies, the Association for Neurogenic Communication Disorders, has an evidence-based treatment set of work groups and practice guidelines and consensus statements along those lines.

MEMBER VAN DE KAMP: As does American Speech and Hearing. There's a number of evidence. And I think if you're -- I mean, I wasn't expecting to have to justify the profession of speech and language pathology.

(Laughter)

MEMBER VAN DE KAMP: Because I've given my career to this whole thing. At this point I think it was well done, but you know.

I think to your point as Rob is saying that if the evidence is something that doesn't meet NQF's requirements to be demonstrated then that's something different than if the evidence supports whether, you know, these services are valuable or not.

And so personally I struggle with our inability to start put forth outcomes. We get caught up in process so frequently that we're almost afraid to judge ourselves by an outcome. And so I want to make sure we don't --

DR. PACE: And that's why -- I mean before we go down this road I really want -- I think we need to have a discussion about this. Because the NQF is really interested in outcomes. Function, health status, it is a huge driving force. Outcomes are integrative of a lot of different care processes and interventions so they're much more efficient than, you know, trying to parse out 20 steps in a process. And we do not require that the

quantity, quality and consistency of a body of evidence be demonstrated for an outcome measure, but that there's some reasonable relationship to services, health care services that are impacting that outcome.

I guess I would think that because we have this whole treatment that is valid enough to refer patients to and to get payment for that there must be some relationship between getting speech-language pathology services and these outcomes. So I just want to try to understand what --

CO-CHAIR TIRSCHWELL: So, can I interrupt for one second?

DR. PACE: Yes.

CO-CHAIR TIRSCHWELL: Let me let Jordan and Salina talk first.

MEMBER EISENSTOCK: So just as a member of the work group I definitely agree that all these measures, this is going to be the sticking point for each of them. And I think it comes down to what Ramon and Therese

were saying which I completely agree with is how to interpret that half of a 2 by 2 situation. And Ramon's point that it doesn't really give us any information or we don't know from any of the data we were able to see what is the natural progression of recovery versus what was the impact from these particular treatments.

I would say that it makes it even more complicated and because there's a chance that we might not get to the measure that I was going to lead, the 0448, in that particular one about memory, and I don't know how we rationalize this or put it all together, 4 hours of treatment had a higher percent making progress than 5-plus hours of treatment. So I think there is some real problems with using this as our NQF-based evidence in that respect. That didn't make much sense.

CO-CHAIR TIRSCHWELL: So, I guess -- and I think you were trying to make this

clarification, Karen. Does the evidence that we're demanding here to get endorsement, does it have to be that we're showing that this measure, there's evidence that this measure itself is driving improvement in outcomes, or do we -- is the evidence that what this measure is addressing which is speech pathology services, is there evidence that that improves outcomes. And if that which is not based on what's here, if that exists then that's sufficient evidence to move forward. Is it the latter?

So they're saying it's the latter. So which seems a little contrary to the vote we had on the first one so I take comments. Jocelyn?

MEMBER J. BAUTISTA: And isn't it still the responsibility of the developer to present that evidence to us? And not for us to do the literature search to find that evidence?

DR. PACE: Right. So this is, you

know, that's what I'm saying. We are not requiring for health outcomes, function being a prime example, that they present a literature review, a body of evidence like we are requiring for process measures. The reason being is that there are multiple processes and interventions that affect these outcomes, not just speech-language but other things that are going on for the patient probably in their initial treatment of the stroke impacts some of these outcomes.

So, we are just saying is there reasonable expectation that health care services -- is there rationale that health care services, in this case speech-language pathology, affects this particular outcome.

CO-CHAIR TIRSCHWELL: And even more did they give us that reasonable expectation in the document --

DR. PACE: Right.

CO-CHAIR TIRSCHWELL: -- that they sent to us. And I guess maybe that's --

DR. PACE: Right.

CO-CHAIR TIRSCHWELL: -- the piece that's missing in all of these.

DR. PACE: And so, you know, I understand what you all are saying about what was presented. And you know, one approach would be if you consider this all insufficient we can put these on the back-burner and continue the work with the developer for a future submission where they can --

CO-CHAIR TIRSCHWELL: Okay. Jolynn?

MEMBER SUKO: I was just going to say unlike the readmission measures which were, you know, driving some interventions, we don't even see interventions of a hypothesis. I mean, you know, we didn't even have reasonable hypothesis about what's driving these based upon what's been submitted. And these have been endorsed for 4 years now, you know, unlike the others. And so I'm just struggling with where we are in terms of our

measure maturation.

CO-CHAIR TIRSCHWELL: Bill?

MEMBER BARSAN: Yes, I'm just -- I guess I'm not really clear this is really an outcome. I mean, I don't know, it's just, it's not clear to me whether this is really more of a process rather than an outcome. I'm just not -- I mean I understand somebody who's alive or dead, that's -- I mean I can separate that and say that's an outcome.

(Laughter)

MEMBER BARSAN: One way or the other, readmission --

CO-CHAIR TIRSCHWELL: That's an ED doc talking.

(Laughter)

MEMBER BARSAN: No, no, seriously. But I'm not sure what the outcome is.

CO-CHAIR TIRSCHWELL: Improvement.

DR. PACE: It's function and it's improvement. It's not just one point in time, it's a change in function and those are

considered outcome measures. We have other -- in other settings we have percent improved in their ADLs, different ADLs as an outcome measure. Those need to be risk-adjusted. So this is, you know, function. These are always more difficult, granted that, but we would classify change in function as an outcome.

CO-CHAIR TIRSCHWELL: Michael?

MEMBER KAPLITT: Don't you need some evidence that it actually makes a difference? Right? I mean, isn't that what we're debating here? We get the point. And your point which is look, you know, we could all get caught up in specific interventions.

I for one am a huge believer, so that nobody's offended, I'm a huge believer in these services, but there's a difference between saying that we inherently believe that they have value and has the level of evidence risen -- has it risen to a level that we're going to hold people to a standard which is what we're talking about I think, unless I'm

misunderstanding what we're doing here.

MEMBER BARRETT: I think I can see where we're all going here and I think that the work group was quite sympathetic to this direction. As we continue potentially down this road I think the work group would probably want to make a couple of comments that modality-specific measures as was said before, outcome measures have been really important to develop and Cramer and a number of other people have written about this. It's to improve the validity of stroke care.

Discipline-specific measures are really important and in particular here we've talked -- yesterday I think when we talked about dysphagia looking at a process that is already endorsed and has value. So before we leave this topic which it sounds like we're moving toward we want to make those comments.

CO-CHAIR TIRSCHWELL: Gail?

MEMBER COONEY: I just worry that the developer was perhaps misled by some of

our verbiage that basically says if you're an outcome measure you don't need to demonstrate evidence. And I would hate to see us throw this out with that kind of misunderstanding.

CO-CHAIR TIRSCHWELL: Jane?

MEMBER SULLIVAN: I want to go back to our work group call and we really struggled with this. And I think A.M. mentioned that one of the pieces of guidance we were given was that evidence was not only what was written but what was clinical judgment. And that's vague. But I also think in terms of the developer the work group really asked for more information about what's happened since the time that this measure was first endorsed and now. And I'm not seeing that we received significant information to further inform our decisions.

CO-CHAIR TIRSCHWELL: Therese, did you have a comment?

MEMBER RICHMOND: Yes, and I understand that we don't need to look at the

evidence like a process measure, but I was on the outcome measures that we talked about all morning so I get it.

(Laughter)

MEMBER RICHMOND: I learned a lot. However, I do think that it was inherent to provide the linkage between structure, process and outcome or one of those, and that is what I believe is missing. So, you know, if -- and I don't know the literatures, but I didn't see the convincing evidence that that linkage was made.

CO-CHAIR TIRSCHWELL: Karen, do you have one final comment or you're good? Well, I guess I would suggest that we vote one more time on the evidence and then depending on the outcome we'll decide what we do with the other measures. So can we go ahead and open up the voting for the evidence for this measure?

MS. THEBERGE: Three yes, eighteen no.

CO-CHAIR TIRSCHWELL: So I guess at this point I would open the floor to the NQF people. It seems like the measure developer, maybe we should check back in, agrees that they're all very similar in this exact regard and would likely receive the same evaluation. So can we somewhat administratively apply the same ruling to the other measures?

DR. MULLEN: The measure developer still feels that way.

DR. BURSTIN: Yes, and this is Helen.

CO-CHAIR TIRSCHWELL: Thank you.

DR. BURSTIN: I think that's quite doable. I do think it's important to look at the list of all of them though and see if there are some in fact where that may not be the case. I mean again, just sitting here trying to read some of the systematic evidence reviews. I'm an internist, not a rehab person. It certainly looks like some of the

issues around swallowing and aphasia might have more evidence than some of the others do.

Again, I think we're still limited by what's in the submission form clearly. So I think it's fine to do, but if anybody wants to just as a process point pull out any other measures for further discussion. Otherwise I think it's fine to proceed.

CO-CHAIR TIRSCHWELL: So for the -- do you want to save an application --

DR. BURSTIN: And have details.

CO-CHAIR TIRSCHWELL: -- before we triage it so to speak?

DR. BURSTIN: Yes.

CO-CHAIR TIRSCHWELL: Anybody want to chime in? Mary.

MEMBER VAN DE KAMP: I guess the only thing I really want to state for public record is that we have to continue to move down the road of outcomes. And I think it is a difficult philosophical discussion in medicine and in rehabilitation, and that I

think is one of the reasons that people have been hesitant to develop these.

And if you look historically at what the government and CMS looks at is they're looking for some more standardized assessment of the care that we provide. And I agree that this application may, you know, the evidence certainly could be -- we could look at broader-based. I mean, we talked very clearly that the assessment of rehabilitation was absolutely critical and we talked about ordering was critical.

Now, okay, we ordered and assessed it but now we're talking about what was really delivered and we have hesitancy to determine if there's an outcome associated with that. And I've seen that over and over again. Everyone's more comfortable with process measures than they are to put the name on the bottom line to say I stand behind that the services I provided changed function. And that's really what these outcomes are

determined to do. I provide --

To your point, I wanted to -- Jordan, to say that's a good statistic that if I gave more than on average 5 and a half hours I didn't get any better than that. But we don't have those kind of measurements. So we think more is better in health care sometimes, so without that measurement.

So I realize that I'm sort of having this little, you know, a bit of a soapbox here but I'm concerned that we don't take this off the table and more comfortably look at well, I did the right process but did it have the right outcome. And that's where I struggle with approving hands down assessment is great yesterday, but do we ask for a lot of basis for why the assessment was so important? And why we ordered? And yet we look at okay, well once we did that then what's the value.

So I understand that you have hesitancy but I just want to be sure that we

don't look beyond the baby steps that we need to do and grow. From the speech pathology side the recognition that we -- why is it only 10 percent? Because no one's forced our industry to step forward to defend through some sort of objective measurement to say what we're doing. And so that's all.

CO-CHAIR TIRSCHWELL: Thank you, Mary. Salina?

MEMBER WADDY: So I don't think any of us as stroke neurologists or practitioners would argue that we will not send our patients for these services. I think the challenge is that we're being asked to vote on evidence that we are not being provided, and we aren't sure of the evidence because that's not necessarily our field in terms of how critically significant the evidence is.

So I'm not clear whether or not it's just that we don't have the evidence for this presentation or if we in general within

health care there's no evidence.

CO-CHAIR TIRSCHWELL: Okay. Risha.

MEMBER GIDWANI: I may be a little confused here, but it seems to me that there are a few different issues floating around, one being the standardization, another potentially being uptake, another being measurement and communication of results as a result of this measure. And I wonder if the fact that this is being endorsed -- I'm sorry, being maintained, it's already been endorsed, shouldn't that already point to the standardization and the need for the evaluation?

If we're putting this measure forth as a means of standardizing the approach I wonder if, (a) has this already been done because it's been in effect for so many years, and (b) if the goal is standardization is this really the appropriate measure to do that, or would another measure that talks about use of

a singular tool across a variety of patients accomplish that.

CO-CHAIR TIRSCHWELL: Okay, thank you. Michael?

MEMBER KAPLITT: Just to reassure you because I think everybody here agrees with what you're saying. But I think there is a distinction though between the assessment and what we're doing here. Because the value of the assessment when you say well, you know, why assess something and then not -- is that if we endorse what we think is a valid assessment tool by enforcing that, by making that a standard that's exactly how we're going to generate the data that will then allow us to determine over time what the appropriate outcome measures should be. So I think there can be a distinction that there can be real value in validating an assessment and sort of trying to encourage that. But I think the outcome measure standard has to rise to a different level.

MEMBER VAN DE KAMP: I agree, although I don't think that the -- if I'm mistaken, I may be mistaken that we didn't have a tool. It still relies on the clinician's judgment. So I get a speech and language order and I use the tools that I feel are most valid. So I think it's the same thing in treatment, it's just that it's harder in looking at health outcomes I think to standardize. So I agree, I understand where we are, I just felt for public record it's important that we move forward with trying to progress this.

CO-CHAIR TIRSCHWELL: And I personally would encourage you not to take this result as an indication that you shouldn't press on with all due effort.

Any other comments before I guess we administratively not endorse. Okay? So then any comments from the developers or -- let's start with developers first.

DR. MULLEN: No comment.

CO-CHAIR TIRSCHWELL: Thank you. And then could we have the operator open up the lines for public comment?

OPERATOR: At this time I would like to remind everyone in order to ask a question press * then the number 1 on your telephone keypad. At this time there are no questions.

CO-CHAIR TIRSCHWELL: Okay. Any comments?

DR. BURSTIN: This is our public comment period for the morning.

MS. TONN: My name is Sarah Tonn and the American Academy of Neurology thanks you for the opportunity to comment for the public record.

The AAN is a medical specialty society representing 25,000 neurologists and neuroscience professionals who have a major stake in providing the highest quality of patient-centered care for stroke which is a neurologic disease.

For the three specific measures addressing in-hospital and 30-day outcome measures which are NQF numbers 0467, 2026 and 2027 the evidence, validity and usability criteria have been endorsed based on rigor of statistical models, yet the models are only as good as the data collected. Administrative data is billing data, it is not data that measures quality of care. The data we need to account for stroke severity, stroke -- transfer issues, patient-centered preference sensitivities, decision-making on comfort care, socioeconomic status and race are missing. For readmissions for stroke care transitions are a huge piece that are not addressed in the model.

The AAN strongly opposes the use of these three measures for public reporting or for use in accountability programs. Endorsement leads to use of these measures in public reporting such as HospitalCompare and this is a disservice to the public as rankings

classified by one vendor method can show higher than expected mortality and lower than expected mortality when classified by one or another method.

A publication by David Shahian and colleagues and they run the thoracic surgery registry, they wrote in the New England Journal of Medicine in 2010 an article on the variability in the measurement of hospital-wide mortality rates comparing vendor methodologies across four vendors.

Each of the four vendors were given the same data. They were given 2.5 million discharges in 83 Massachusetts hospitals over a 3-year period. All these vendors were given that same data. Four vendors, UHC which is University Health Consortium, 3M, Thompson Reuters, Foster which is out of Imperial College London, when comparing vendor methods the findings were that a total of 12 of 28, a little less than half that had higher than -- of 12 of 28

hospitals had higher than expected mortality rates when classified by one method, and yet had lower than expected mortality when classified by one or more of the other methods.

Addressing one more point in particular also supports the AAN's strong opposition to endorsement of these three outcome measures. The strongest predictor of short-term outcomes among stroke patients is baseline stroke severity.

The baseline NIHSS or National Institute of Health Stroke Scale score has more predictive power than all other baseline variables, demographics, comorbidities, et cetera, combined. Therefore, evaluating short-term outcomes without adjusting for baseline stroke severity will always be subject to missing variable bias.

Smith and colleagues in the Circulation 2009 publication demonstrated in prediction of in-hospital mortality in

ischemic stroke using data from Get With the Guidelines that using NIHSS alone produced a C statistic of 0.83. Imagine if all the C statistics reported today in the models adjusting for less important factors has this impact realized the missed opportunity by not adjusting for stroke severity.

More scientifically sound and rigorous approach would be to collect the needed data and subsequently use it to adjust and validate the in-hospital and 30-day outcome measures. If the appropriate data is not collected and compared to the in-hospital and 30-day outcome quality measures then it will be impossible to accurately assess quality of care and likely will significantly penalize the tertiary care centers.

The AAN's opposition is expressed in the AAN 2010 letter to CMS included in your steering committee materials. Thank you.

CO-CHAIR TIRSCHWELL: Thank you very much. Any other comments? No. Okay.

Let's see. What is next on our agenda? Lunch. And then what's left for this afternoon, can you guys review with us? Just the related and competing measure evaluations.

DR. PACE: Are you caught up on time now?

CO-CHAIR TIRSCHWELL: Well, we're half an hour behind. So why don't we take a 20-minute lunch, a little bit short, try to reconvene at 12:50 and maybe we can get it all done.

(Whereupon, the foregoing matter went off the record at 12:31 p.m. and resumed at 12:53 p.m.)

CO-CHAIR TIRSCHWELL: Okay, we're going to get started again and Karen's going to give us an introduction to what we need to do with the next phase here.

MS. JOHNSON: And Suzanne is bringing up our slides. We talked about this a little bit yesterday afternoon. I gave you the birds' eye view. Today we're going to go

through a little bit more detail about the related and competing.

Okay, just to remind you NQF does ask you to consider issues of related and competing measures. So if a measure meets the four criteria which that's what you've done in your meeting so far and there are endorsed or new measures that are related. So related we define as having the same measure focus or the same target population, or if there are competing measures which we define as having the same measure focus and the same target population then we ask you to compare them to address harmonization or selection of the best measure.

So if you are looking at related measures we want you to evaluate whether the measures are harmonized, and by that we mean aligned as much as possible in terms of their specifications, or if they're not are the differences justified.

For competing measures we ask you

to compare and if possible choose a superior measure. And if you can't choose a superior measure then put forward some reasons why multiple measures would be justified.

Okay, and pretty much this is in chart form what we just said. So again, what you are thinking about here in terms of competing versus related is how the numerators and denominators basically match up.

And again, to remind you of what we said yesterday we're not talking about is the measure focus exactly the same. We're asking are they conceptually the same because there's going to be little, tiny differences amongst these measures. And again as we said yesterday we have identified those for you that we would ask you to look at today.

CO-CHAIR TIRSCHWELL: Can I ask a question about that? So, there's the different target patient population but what about the different target evaluation level, like the clinician ones versus the facility

ones?

MS. JOHNSON: Those we actually consider. Conceptually those would be, all other things being the same, we would consider those to be competing measures.

However, there often may be reasons why you think that having the two different levels of analysis is important. So in that case you would say I cannot pick a superior measure and here's one of the reasons. We think it's important to have a facility-level measure and a clinician-level measure. Am I saying that correctly, Helen?

DR. BURSTIN: Yes. The only thing I'd add is that in that case the most important issue is really harmonization. So for example, if you were talking, I think one of your measures you're going to talk about is some of the DVT prophylaxis work, clinician-level versus hospital-level.

Granted the data systems are so completely different in this day and age it's

almost impossible to get a measure that will reflect both if you think both are important, and in that case you would want to make sure that the clinician-level assessment matches the hospital-level assessment and is harmonized on the most important data elements.

CO-CHAIR TIRSCHWELL: And it would seem to me that one of the reasons -- predicting why you might need a clinician-level one and a facility-level one is that there are these giant processes that are rolling out that are going to be evaluating clinicians on these levels and are already evaluating facilities on these levels. And it sort of seems you're, I don't know, you probably need to endorse them both. Gail?

MEMBER COONEY: But wouldn't it be cool if we had a way to like know when we were looking at the same patient from the facility assessment and the clinician assessment and see whether we got the same conclusions?

DR. BURSTIN: So that really speaks to I think what is the ideal state, right? Wouldn't you love to have the measure that rolls up and rolls down in terms of higher levels of aggregation and down to the clinician as appropriate. That's where I think we all want to go.

I think just given where we are in terms of data systems in America at this point until we sort of get to some better data infrastructure, particularly EHRs we hope, that's harder to do.

MS. JOHNSON: Okay, and as we go through this exercise in a few minutes once we talk about -- just to kind of emphasize what Helen said. Once we talk about competing measures we will then ask you, if you say that the differences are justified we will ask you a little bit more about harmonization. So that's kind of the second part of what we do.

So what we have here in the next three slides are the thinking steps that we go

through to see whether or not things would be recommended.

So, the first one here before we even get to related and competing is does the measure meet all four of the NQF evaluation criteria. If you have said no then we do not recommend it and we're done, and that's what you guys have done this morning. If yes, then are there potentially related or competing endorsed or new measures? And if yes, then we ask you to compare the specifications.

And at the conceptual level do the measures address either the same measure focus or the same target population. If so, if it's the same measure focus but a different patient population we ask is there a way to combine the measures. And if so, then we would say recommend the measure that has combined those populations. Does that make sense? Hopefully that makes sense.

If those measures can't be combined in some way then we go down to the

next level and if the measures address the same concepts for a measure focus for the same population we call those competing measures. And then what we do is we ask you to compare the specs and basically we go through this little exercise here.

If they're very similar we ask the measure developers can they resolve the stewardship so that they can create one measure. That is a little more difficult if the two different measures are from to different developers obviously. Sometimes the answer's yes and sometimes the answer is no. If no, then you go on and compare both measures on all of the evaluation criteria and weigh the strengths and weaknesses across the criteria and try to determine if you can whether or not one measure is superior over the other.

So for example, if one measure, the validity you thought was iffy so the vote was really close on validity on one, but you

were very happy with validity on another, then that might be a reason that you would think one is superior to another. That's just an example.

Again, if you cannot recommend one of those measures as superior then we would ask you to say is there a justification for having multiple measures. And if so, then what you would be doing is putting forward your recommendation that both measures would be put forward. Otherwise you would -- and if you do think one is superior then obviously the one that you think is superior is the one that would go forward. The one that you think is not superior would go down at this point. So it's very possible theoretically that something that you thumbs-upped yesterday could go down today because you think it's just not quite as good as another one that's very similar. Okay. And then if we can go to the next slide.

Related measures we don't ask you

to choose the superior one because they are going to be different either on the measure focus or on the target population. But we ask you to look at the specifications and see whether or not they're completely harmonized. And what's it say. Compare -- are they completely harmonized? Yes. Recommend one measure. I think technically that would be -- you would recommend both of those. Okay.

If they're not completely harmonized then we ask you to consider are there reasons that there are differences in the specifications of the measures.

DR. PACE: I think what this is going through -- sorry. I think we have a -- it's my slide. I'm sorry, I think there's a confusion here. What's the next slide? Could you go to that? Okay.

This is still -- and then go back. So, I think what we're -- this is actually if you're still talking about competing measures and you've decided that you have to move

forward with both measures then you still want some harmonization to the extent possible. So, if you have two measures of -- well, we do have two somewhat competing measures of hospital-level mortality. There are some differences but essentially they're trying to get at measuring mortality. You may decide that yes, we need both of those measures, the 30-day and the inpatient, and then we would want them harmonized to the extent possible. So, we can just go on then.

MS. JOHNSON: And I guess where -- I'm a little lost here, Karen. Help me out.

DR. PACE: So if you -- you would want to send the -- we'll just move on past this slide. The idea is that we would ask the measure developers to look at opportunities for harmonization. And you may have specific recommendations. For example, the definition of a readmission should be the same across both measures.

MS. JOHNSON: So when you're

thinking about superiority, I think we've gone through this but pretty much we want you to think about impact, opportunity and evidence. So that's criterion number one, importance to measure and report. Think about the scientific acceptability.

So here, this is a few more examples. Untested measures cannot be considered superior. In this phase we do not have any untested measures, but in phase 2 it's actually likely that we may be seeing some of those untested measures.

And then we would also have a preference for measures with the broadest application as well as those that address disparities in care. So again, I think yesterday I used the example of two measures, one that looks at patients 18 and older versus one that looks at 65 and older. We would prefer one that has the more broad applicability.

For usability we would prefer

measures and ask you to consider as superior those measures that are actually being publicly reported or in widest use or even in use as opposed to perhaps just being planned for use.

And then feasibility, obviously measures that are based on electronic sources versus ones that would require more manual extraction, for example, might be considered superior. Freely available. This speaks to the mortality measure. We saw one this morning that was based on the 3M algorithm for the diagnosis classifications.

DR. PACE: However, that's available.

MS. JOHNSON: It is available. Yes, we found out this morning that we could actually go in and look at that. But that would be an example of what we used to think wasn't freely available. Next slide, please.

Justification of multiple measures. Basically what we're asking here is

to assess the value versus burden. So, what's the value of having more than one measure compared to the burden? The value is perhaps there is an EHR-based measure that's very similar to one that's paper-based and there's good reason that you would want to bring aboard an EHR-based measure as well. But maybe there's measures with broader applicability but still can't bring in every patient population, every setting, that sort of thing, as well as increased availability of performance results. So, this may speak to measures that have been around for awhile longer.

Burden, again, are things like increased data collection. Is it worth having to collect data two or more times for two different measures? And if you have multiple measures do they give you similar results across. Can you interpret them similarly? And if not, then that's kind of a burden of having multiple measures. So again you would

decide what you think about whether the value of having multiple measures outweighs the burden. And again, that's your expertise and judgment that we're asking you to use here.

For lack of harmonization, again that refers to having two measures that are related but perhaps the specifications are different. And again there may be justifications for having different specifications. So, what you want to do is think about the evidence. Is there evidence for one specification versus another. Remember that different data sources may require some differences in technical specifications. So if you were having a measure that's based on claims versus one that's based on a paper abstraction they might have to give different details of how to abstract that data. They should not be simply due to proprietary interests or preferences.

The difference -- we've already addressed this. The difference would not

affect interpretability or burden of data collection. And again, if it does affect burden then you would decide as you would with the competing measures for superiority does value outweigh the burden. Okay, so next slide.

Before we go onto the meat of the discussion let me open it up for questions and see what you think. Terry's laughing so something's wrong.

MEMBER RICHMOND: I'm just thinking I don't really have all that, but I think going through it may be helpful. When we do the measures I'll get it.

MS. JOHNSON: What we're going to do today -- Risha?

MEMBER GIDWANI: I have just a specific question. I'm trying to find the 3M APR-DRG grouper online. So does anybody have an actual link for that? The 3M website is not proving fruitful for me.

CO-CHAIR TIRSCHWELL: The one

where they said all the details should be revealed behind the magic curtain?

MEMBER GIDWANI: Do you need a password for it?

CO-CHAIR TIRSCHWELL: My guess is it's purposely hard to find, but --

MEMBER RICHMOND: Actually I think they said that you need a password.

CO-CHAIR TIRSCHWELL: You probably have to sign up, give your email address, your firstborn.

DR. PACE: We'll get that in the steering committee.

MR. GEPPERT: I'm sorry, this is Jeff from AHRQ. I can give you the URL if you want it.

MEMBER GIDWANI: Yes, please.

MR. GEPPERT: So it's www.aprdrgassign -- A-S-S-I-G-N -- .com. And you do need a username and a password but hopefully there are instructions there.

MEMBER GIDWANI: If you don't

already have a login and there is no place to register for one is there a basic login you can give me?

MR. GEPPERT: We just have to use the login information that they gave AHRQ. I can try to email them quickly and see if they can provide a guest password.

MEMBER GIDWANI: Okay. I just can't register for one so that would be great.

CO-CHAIR TIRSCHWELL: So that's probably not going to affect this so if you'll move that offline for the moment.

CO-CHAIR KNOWLTON: Karen, I have a question. Just -- it's not related to what you said up there, but I noticed that in the material you gave us that some of the things that we are considering we did not consider as a steering committee. How does that work?

MS. JOHNSON: Okay, how that works is there may be measures, and there are actually, that you guys did not consider in this project but they have been considered in

other projects and are currently endorsed. So we would still have you look at the specifications. Obviously you will not have gone through the entire thinking process that you did with the measures that you looked at in this project, but as best you can does one look superior or is there room for harmonization.

Basically if you decided, if you had a competing measure from a different project and you decided that the one that you looked at in this project was not the superior measure then what that would do is that would take that one down. Okay?

If you decided that the other one based on your brief review was not the superior measure then really you don't have any control over that. It would not affect endorsement at all. It would just be your information that you think that the one that you looked at in this project is a better one than the one that's out there already.

And let me ask Helen and Karen and make sure I have that correct.

DR. PACE: Yes. Right now all you can do is act on the measures that are in this project. But we would like your recommendation because that's something that would then come up when that other measure is up for endorsement again.

MS. JOHNSON: Any other questions before we delve in? Okay.

If you have your homework from last night that we passed out, and I don't really expect you guys to have looked at these in detail last night, but if you would pull that up and go to the last set. I think it should be group number five. We're going out of order here because someone from the developer team needs to leave a little earlier so we're going to do the mortality measures first. Okay.

CO-CHAIR TIRSCHWELL: Page 13 I think in the handout.

MS. JOHNSON: Page 12 and 13? Oh. It's mislabeled. It should be 5 I think. Sorry about that. It's the last two pages.

Okay, so what I've tried to do with this handout is to give you some really basic information about the measures. So I've given you the description, the numerator, denominator, the exclusions, and then I've actually gone through and told you what the measure focus is and the patient population. I've given you the time frame, setting and level of analysis. And also data source.

And then we also put in -- actually, we don't have these. Are these the new ones? Jessica is printing out the new ones now. What we -- we have a new version of this because when we handed this out last night we had not done the stroke mortality measures so we couldn't tell you what you guys had voted. So we have more coming to you. But I don't know necessarily that for beginning our discussion if you have to have

that. Let's just walk through and see if we can, see how far we get.

Let's think about the mortality measures. We have one from AHRQ and one from CMS. And since both are measuring mortality and since both are measuring mortality in stroke patients we consider them competing measures. Okay? So since we're thinking of them as competing measures we would like you to think about whether or not you would consider one of them superior. Salina.

MEMBER WADDY: 0467, isn't that ischemic stroke and hemorrhages? Whereas the others are just ischemic stroke?

MS. JOHNSON: You are correct. So --

MEMBER WADDY: Do you all have some other measure that's already ongoing for hemorrhages? Other than this? Okay.

MS. JOHNSON: Operator, can you mute those lines?

MEMBER J. BAUTISTA: I also think,

I mean inpatient mortality is very different than 30-day mortality. So I have trouble thinking of these as competing measures. I think they're very different.

MS. JOHNSON: Again, we are considering them competing because the measure focus is mortality. So again it's a very conceptual, very high-level way of thinking about it. But it could very well be that in your mind it's important to have an in-hospital measure as well as 30-day longer outlook. In which case you would -- that would be one reason that you might say multiple measures are justified.

DR. PACE: So let me -- can I just add to that? So we start out with, you know, the broad concepts, what it's trying to measure, and then there may be very important reasons that we should have an inpatient and a 30-day and that's what we want you all to weigh in on. So it's not that we're saying that you will have to choose between them, but

we want to start with that discussion and then go from there in terms of if there are harmonization issues.

We prefer to do that then starting with because there are differences that we automatically accept that there should be differences. So it's not that you're going to have to but we want to at least introduce that question for you to work through.

DR. BURSTIN: And just to give you an example, for AMI which is actually very similar we actually -- the cardiovascular committee decided both inpatient and 30-day were related but both important concepts.

MS. JOHNSON: I think David.

CO-CHAIR TIRSCHWELL: So I was just going to comment that although they're competing by that loosest of definitions I think there are important differences which probably do justify it. And just a couple of the big ones are the inpatient versus 30-day. That's a really big difference. And I prefer

the 30-day one in that scenario.

On the other hand, the inpatient one has all ages and all stroke types whereas the 30-day one is restricted in age to greater than 65 and only ischemic stroke. So in that one I actually prefer the inpatient one. So I think there are sort of arguments on both sides is my particular perspective.

MS. JOHNSON: Bill?

MEMBER BARSAN: Yes, so is there ever the possibility on things like this -- I mean, you know, what would be great is would be to have one measure that looks at anybody from 18 and over. You could split it out with 65 and over and then, you know, all the rest of the adults. You could split out ischemic stroke, you could split out hemorrhagic stroke. But you just have one measure but you just specify that it examine all those different subtypes. That would be ideal rather than having three different measures.

CO-CHAIR TIRSCHWELL: Can I just

respond back to that? I mean the methodologies of these are so different that that would be a totally new measure.

MS. JOHNSON: Karen?

DR. PACE: I mean obviously that's also a preference for NQF. Measures that have the broadest applicability capture the widest target population indicated by the measure or the evidence that you could do those kinds of things, you know, inpatient and 30-day.

Unfortunately, you know, that's a goal in the future and not our current realities. But I think you could ask the measure developer -- I don't know if these particular measure developers have discussed combining forces in any way to move toward that goal.

MEMBER R. BAUTISTA: If I have two measures, one is a hospital measure, the other one being a clinician measure, and decide to take the clinician but not the hospital measure basically at the end, are we in fact

saying that we're no longer going to be judging the hospital for example? For that particular measure. Or is it going to apply for both clinician and hospital at the end?

MS. JOHNSON: Just to clarify, both of these are going to be hospital measures that we're talking about right now in group 5. But your question is more generally if you had one versus the other. If you did choose one to be superior then by definition you would be saying that the other one you would not continue to recommend for endorsement. So the other one would go down.

MEMBER R. BAUTISTA: Right. You're basically you're not going to endorse. It's as good as non-endorsement then I suppose, right?

MS. JOHNSON: Yes.

MEMBER R. BAUTISTA: Okay.

MEMBER COONEY: But couldn't we at least like under the description one measures both hemorrhagic and ischemic stroke, one

measures just ischemic stroke. Can we get them to agree to both measure the same thing there?

MS. JOHNSON: We could certainly ask them to respond to that.

PARTICIPANT: Hi, this is -- from Yale. I don't know, Susannah, are you still in the room? Or she had to go catch a train. If not I'll speak to that. Can you hear me?

DR. BURSTIN: She's here, are you Elizabeth? And Susannah is still here as well. So she's coming up.

PARTICIPANT: Okay, great. I defer to her then.

DR. BURSTIN: Okay.

DR. BERNHEIM: So I think there may be value to looking at hemorrhagic stroke as well. The clinician group that supported our measure development felt very strongly that the combined was too heterogenous a group to adequately produce a good 30-day risk adjustment model. So I think we would need to

have conversations and the question really might be about splitting out and developing a second measure. But I think it would be hard for our technical expert panel to swallow a combined measure.

PARTICIPANT: I just want to point out that the death rate for hemorrhagic stroke is just so much higher as well.

MS. JOHNSON: David?

DR. ROMANO: This is Dr. Romano, could I address that?

CO-CHAIR TIRSCHWELL: Sorry, go ahead.

DR. ROMANO: So we both have had -- internal discussion of this issue both with the analytic group as well as expert panels. And we generally take the perspective that these cohorts for analysis of risk adjustment mortality should be defined based on characteristics of the patient that are apparent before admission, presentation to the hospital.

This is a very similar situation that we have for example with heart attack where we have FT elevation MI versus non-FT elevation MI. Two pathophysiologically different conditions with different outcomes but patients don't know which one they have and doctors don't know which one they have before the patient arrives at the hospital.

Similarly, for heart failure we have systolic and diastolic heart failure combined. So in general for pneumonia we have viral and bacterial pneumonia combined. So in general we prefer to define the cohorts based on clinical presentation that are clear in -- before patients arrive at the hospital or when they first arrive in the emergency department so that our analysis is not susceptible to bias by the diagnostic process, or less susceptible to bias by the diagnostic process.

But having said that we have as Jeff Geppert alluded to earlier estimated stratified models for ischemic and hemorrhagic

strokes, and we would certainly be willing to consider. We have heard feedback from the user community that they would like to see separate mortality estimates for ischemic and hemorrhagic mortality. So we would be willing to consider stratified model with stratified reporting with a composite measure then being the primary overall outcome.

MS. JOHNSON: Michael.

MEMBER KAPLITT: So based on that logic why not include patients who come in with transient aphasia and then it turns out they had a seizure from an undiagnosed brain tumor that was only found out after they came in? I mean, if you're going to base it purely on their clinical presentation, you know, based on that logic then you know, you would include a lot of things that aren't even strokes. So obviously there's something about the diagnostic process that winds up influencing this.

I think the discussion that we all

had this morning as people who treat these things for a living is that we feel that there's such a fundamental difference between a hemorrhagic stroke and an ischemic stroke that that does, you know, warrant potentially special consideration. That's why we spent so much time talking about it I think.

MS. JOHNSON: Dan? I'm sorry, go ahead, Patrick.

DR. ROMANO: Oh, I was going to say I think the point's well taken. It's a gray line and certainly for other conditions as well. The presentation of pneumonia is not always clear, the presentation of heart failure is not always clear. But -- and we're trying to define -- we're trying to define the right point on this spectrum to define a clinical syndrome that has high mortality where there's opportunities for intervention to improve outcomes and where it's a recognized entity by both patients and clinicians. So, we acknowledge that you could

draw the line in different places and certainly, as I said, we would be willing to and have done some initial analyses to support a stratified analysis.

MS. JOHNSON: Dan?

MEMBER LABOVITZ: I'm sorry to be confused. I'd like to present myself as being smart but I'm feeling dumb.

I don't see how we can harmonize or relate these measures. They are about death which gives me deep anxiety but I guess I already said that. But beyond that they have radically different engines underneath.

They just don't measure the same thing, they really don't. The adjustments are completely different, the data are different. I don't see how you could -- if you could make them look the same on this piece of paper but they are always, always going to be radically different. So what's the point?

DR. PACE: It's an appropriate question and I guess the issue is, you know,

whether you all see any opportunities for harmonization or not, or recommendations for the future. Are there any ways to look at these outgoing because they will come back for endorsement maintenance. Perhaps not. But you're absolutely right, you know, the harmonization especially when you get into these complex measures is much more complicated than when you're talking about a process measure.

MS. JOHNSON: Ramon, I think you were next.

MEMBER R. BAUTISTA: I still don't see why we can compare a hospital-based and a practitioner-based measure. And then try to pick one, it seems to be a little bit worse compared to the other one. In other words, we're actually going to be not endorsing something that we endorsed before because it's a little bit worse compared to another measure that's meant for a totally different target population. I don't understand that. It's

probably a backhanded way of trying to get rid of something you didn't like in the first place.

MS. JOHNSON: We will table that question until we get to a set of measures where one is facility and one is clinician. Right now both of these are facility-level measures. So we'll come back to your question.

DR. PACE: I might just make one comment though. At that level it's probably going to be about harmonization, not choosing one over the other.

MS. JOHNSON: Terry?

MEMBER RICHMOND: At this point I would advocate keeping both. And I mean, other than one is ischemic and hemorrhagic and the other is ischemic only the methods are so different that my understanding with the CMS measure, the 30-day measure is we're using historical data that are available in patients over 65 years of age which is not available in

this broader population of 18 years and older. And certainly with hemorrhagic and ischemic we will see patients between 18 and 65 that we would lose if we tried to bring those together.

MS. JOHNSON: Did you have any additional? Okay. Then David has to talk.

CO-CHAIR TIRSCHWELL: So I completely agree and I guess the -- I would move to suggest because I don't know that we have the power to do anything else that a little bit of harmonization would be to ask the AHRQ folks if they would really move forward with the subtype-specific models which would allow a little bit of comparison granted it's not the same thing. And I would also implore them to divide stroke into three categories, ischemic stroke, intracerebral hemorrhage and subarachnoid hemorrhage which are all quite different. And it's the classic division of stroke subtypes. And thank God, division by ICD-9 codes as well, so.

MS. JOHNSON: Any other ideas for potential harmonization? Risha. I'm sorry, Risha?

MEMBER GIDWANI: I just want to say I agree with Daniel. I think that harmonization is actually not going to work very well in this case because the data elements are extremely different. And I don't see how they could be reconciled.

I would also love to be able to compare the two models side by side in terms of their coefficients, the variables they're using for risk adjustment and to look at the direction of the effect. Because when I look at the AHRQ models, when I look at the predictor variables, the direction of the effect in terms of whether something is protective or not against mortality has good face validity. When I look at the CMS models and I see things -- gosh, I have too many papers in front of me. But you know, I see that there's not very good face validity and

there are clinical conditions that are protective against mortality -- here we go -- such as heart disease and aneurysm and hypertension, you know, that leads me to not only believe that harmonization is not going to be possible but it puts my balance of favor towards the AHRQ models if we do need to choose one over the other.

MS. JOHNSON: Salina, did you -- was your question resolved?

MEMBER WADDY: Pretty much. I just -- the last -- I completely agree with the proposal that or recommendation or whatever is currently on the floor. My only question is about the age groups at this point. Are we -- what are the age groups that are under consideration for these? Is it over 18 and then the split with the greater than 65, or is it 18 and up for everything?

MS. JOHNSON: Well, as they stand right now the AHRQ measure is 18-plus and the CMS measure is only 65-plus. So I guess --

DR. BURSTIN: And just to speak to that. This has been an issue. Obviously these measures were developed for Medicare to have the data that allows the cross-links between hospitals.

In the past CMS and Yale have done additional analyses to demonstrate whether the risk model itself works for patients under 65 so it could be, for example, used for other patients and settings. So that might be a recommendation if you'd like the committee to take a look at it.

MEMBER WADDY: Yes, I mean I think it at least should -- I think it would be very helpful to have over 18, but then I'm also very concerned, and this may be beyond what we do here, that you know, we could be excluding a really important population which is care in pediatric patients. And so I don't know if you all have a pediatric care section that's separate and if I'm just opening a giant can of worms that maybe can be tabled for a later

time. But at least that --

DR. BURSTIN: Small can of worms. No, it's actually a good question. We have tried to eliminate any setting-specific or population-specific committees because we get into these issues of people being left out. We have intentionally put a pediatric neurologist on this committee to really kind of keep us honest and say if there's measures that come forward and go there's really no reason this measure is X age and up, for example. Isn't it --

MS. JOHNSON: I think it's Dr. Sheth.

DR. BURSTIN: Who just left, yes. So I mean I think it's fair game to indicate if there's any of these measures that we're looking at that are applicable more broadly it's a good question for the developers.

MEMBER COONEY: I just was looking at the exclusions and I don't see why they can't be a little more consistent as well. I

mean, I know we can't completely harmonize them and that they're different, but as much as we can bring into synch as we can it just seems we should.

DR. PACE: So, I think one of the things that in terms of what we've heard so far is that you all are thinking that the value of having these two measures because neither one can accommodate either all of the patients and age or all of the strokes or a variety of, you know, inpatient and 30-days. That at this point in time you're saying that the value of having those two measures is there, and that perhaps you want the developers to look at whether they can do any harmonization with -- AHRQ is going to look at whether they can do a stratification. And the question that just came up is whether there's any room for some alignment with the exclusions that we could just ask the developers to give you a response on that.

Is there anything else that anyone

wants to?

MEMBER WADDY: Just one more thing. I think these were the two measures that had the differences in transfers, how they handle the transfers. So they may need to really look at that.

DR. ROMANO: This is Dr. Romano, can I address that?

MS. JOHNSON: Go ahead.

DR. ROMANO: Yes, I mean I would say that for each of these two measures the exclusions are the correct exclusions for the type of data that are being used. So the CMS measure excludes transfers that come in from another acute care hospital because a death that occurs after transfer is attributed back to the hospital where the patient was originally admitted and that's absolutely the correct thing to do in a 30-day measure.

In an inpatient mortality measure it's obviously not possible to make that attribution. So, we're forced to exclude --

to avoid double-counting we're forced to exclude the patients who are transferred out, because when the patients transferred out then we lose the information about their ultimate outcome when they're discharged from the acute care hospital. So I think that difference is inextricably linked with the difference in source data.

Otherwise our exclusions for children and pregnancy are of course because those are very peculiar populations with respect to stroke. And for pregnancy there are specific coding issues that make it much more complicated to identify strokes and risk factors for strokes. For children of course strokes occur in the setting of very specific high-risk chronic diseases that really make it an entirely different clinical issue.

MEMBER WADDY: I do agree in our previous discussion for I believe it was 2026 regarding how you handled transfers, but I don't think that there is necessarily, at

least not agreement from me regarding how transfers were handled in 0467, and whether or not that's just a limitation of the data set or there needs to be reconsideration of how that data is collected. But it's an important piece of information that I don't think should be glossed over.

CO-CHAIR TIRSCHWELL: I do have something related and that's that in the AHRQ which are the administrative data which are based on that UB-40 form with the diagnosis codes I believe there's a field for admission source which includes the possibility of transferred in from another hospital. Am I wrong about that, Patrick?

DR. ROMANO: Yes, point of origin. And like I say, we do test that in all our risk adjustment models. And it was not significant for stroke mortality.

CO-CHAIR TIRSCHWELL: So it's not that you couldn't exclude them, it's that you decided not to.

DR. ROMANO: Correct. Our usual approach is to adjust for it. So if it reflects a more severe cohort of patients then we want to take that into consideration in the risk adjustment. In this case what we effectively see, and we see this in some other cohorts as well, is that some patients are transferred after they survive the highest risk period, and they're transferred for the purpose of getting additional diagnostic testing, maybe rehabilitation. Other patients are transferred because they're getting worse and they're beyond the capacity of the original hospital to manage. So, there's a washout essentially and that's why those transfers don't show up as having higher mortality.

MS. JOHNSON: Gwen?

MEMBER BUHR: So I was just wondering if it would be possible for the AHRQ measure to also stratify by 65 and older and the younger ages so that they could be

somewhat comparable as well.

DR. ROMANO: Well, in general of course there's a danger to multi-way stratification because that leads to smaller cell sizes and less reliable estimates. But you know, we're certainly willing to hear different opinions. In general our expert committees and our stakeholders have favored including all adults.

MS. JOHNSON: Risha?

MEMBER GIDWANI: I think given the vast number of strokes that occur in the United States every year the cell sizes would be sufficient even if you do stratify 65 and above and lower than 65.

I think given the comments by the American Academy of Neurology indicating that different risk adjustment models will rank hospitals relatively differently using, you know, depending on their own risk adjustment methodology this would be a really great way for when these measures come up for

maintenance to get some data about how AHRQ is evaluating hospitals with patients above 65 versus how CMS is.

And even noting the difference in 30-day versus in-hospital mortality on would assume that the direction of effect of those measures would be in the same direction. And so I think just in terms of the fact that the field of risk adjustment is relatively new this would provide some valuable feedback as to the ability of risk adjustment to rank hospitals in a consistent manner.

DR. BOTT: This is John Bott with AHRQ. This is a question more so for Jeff, but I believe stratifying by age is one of the, what we call canned stratifiers within virtually all or all our QIs. Is that correct, Jeff? So the person currently has the ability to stratify by age, not just 65 and up but by various ways.

MR. GEPPERT: Yes, that's correct. That's a feature of the software that

implements the AHRQ measures, stratification by age.

MS. JOHNSON: Does anyone else have any other ideas for harmonization or have we pretty much covered them? Risha?

MEMBER GIDWANI: I just have a follow-up question to that response. Does that mean that the expected mortality is also stratified by age, or is that just the observed mortality that you can stratify by age?

MR. GEPPERT: It's both. So the difference I think between just simply reporting the rates stratified by age and developing a separate measure that's stratified by age is you would develop a separate risk model for the strata. So the data is slightly different.

MEMBER GIDWANI: You're saying -- wouldn't the risk model be the same? It would just include a data set that had patients 65 and older.

MR. GEPPERT: I'm saying that if -- there's two options. You can either just stratify the rates by age. In that case it's the same risk model and you simply stratify the rates, divide the patient population into two. But the suggestion of creating a separate measure that is stratified by age you would actually develop a separate risk model for that strata. You have a separate set of coefficients and you would estimate it separately.

MEMBER GIDWANI: Okay, thank you.

MS. JOHNSON: And Karen and Helen, I'm unclear. Is there any voting in this section?

DR. PACE: Not necessarily. So I mean, unless anyone disagrees with what has been brought up before in terms of is it -- if anyone wants to call the question certainly we can. But the -- what we're hearing is that people see value in having both of these measures. So is there anyone who wanted to

speak to the contrary? Okay. All right.

MS. JOHNSON: Okay, great. We've gotten through one of our many that we wanted to discuss. The second one is probably very easy. I put it up here more for completion than anything, but that is actually thinking about the two CMS measures, one for mortality, one for readmission. Those are related measures and it might be at this point easier just to ask the developers are these measures harmonized in order to your definitions and such. Obviously your risk model is different but the approach is the same.

PARTICIPANT: So the measures are actually harmonized. There are some differences as we discussed earlier, particularly around transfer patients. So the readmissions. In a case where the patient is transferred the mortality would be attributed to the hospital that first admitted the patient and the readmission is attributed to the hospital discharging the patient to the

non-acute setting. So in those cases that would be different.

For the mortality measure we randomly select one hospitalization a year if a patient has multiple hospitalizations and for the readmission measure we don't do that but we do block out any admissions that occur within 30 days of an index admission so that no hospitalization would be considered both an index admission and a readmission.

Obviously the risk adjustment variables that feed into the model are slightly different because they're different outcomes. But in other ways they are aligned sometimes.

DR. PACE: So I don't think there's really any issues that -- unless anyone in the committee has identified anything.

MS. JOHNSON: Okay, let's go back to measure group number 1. So flip back to the beginning of your handout.

Okay, just to orient everybody this is the set of measures that looked at antithrombotic therapy. Two measures looked at discharge. So was antithrombotic therapy ordered at discharge or prescribed at discharge. And then another one, was it done by end of hospital day two.

So if you compare the two -- if you compare the two Joint Commission measures, one looking at therapy by discharge, the other looking at therapy by hospital day two, we could consider these either related or competing. And I guess our question was was there any feeling from the committee that these might be combined in some way or is it necessary to have two measures. If we can just get some discussion on that.

CO-CHAIR TIRSCHWELL: I guess I would throw out to the developer who I guess is the Joint Commission as to -- because both of these measures were also kind of topped out or pretty close to topped out. At the patient

level, and maybe they don't have this information.

It might be something they could query the Get With the Guidelines people, but at the patient level if you always achieve the day two and then you always achieve the discharge one maybe you don't need the discharge one if you've gotten the day two. So you know, where is there more room for improvement? If they're, you know, almost 100 percent linked we might be able to collapse into one, but I think somebody would have to sort of analyze some data. And again, it's probably best done from Get With the Guidelines. They might want to work with them to see if they could figure out that type of thing.

MS. WATT: This is Ann with the Joint Commission. Karen and I are both here and we're both going to answer. The answer to your question, yes, we do have the capacity to look at relative results for the same patient

for both measures and that's a good suggestion. It's something we can do. But we're not entirely certain that these are -- that they have the same focus. And I'll let Karen address that.

MS. KOLBUSZ: I think that how we see the focus is in the management of the patient. And stroke 5, which is antithrombotic by end of day two, is really looking at early management of the patient, giving that aspirin as soon as possible after arrival. Whereas stroke 2 which is discharge on antithrombotic is looking at secondary prevention, long-term antithrombotic therapy. So the focus is very different in our opinion.

CO-CHAIR TIRSCHWELL: Just to reply to that, I would think -- it's a good point. I think that you would -- probably wouldn't get rid of the day two one, but it's possible in my mind that you might be able to get rid of the discharge one if they're just continuing along.

MEMBER WADDY: I also think -- weren't there differences as well between these two in terms of the inclusion of TIAs?

MS. JOHNSON: We're looking at 0435, 0438.

MEMBER WADDY: Right, no --

MS. JOHNSON: So look at the second and third columns of that sheet first. We're doing those two first.

MEMBER WADDY: Oh, okay. Okay, got it.

MS. JOHNSON: Jocelyn.

MEMBER J. BAUTISTA: So I really like the idea of combining the two measures similar to what they've done with some of the core measures, the appropriateness of care, composite measure. What percentage of patients receive appropriate care at both time points. If we've topped out at each one then we should move to the next level of, you know, what we expect of patients, of providers, that you have appropriateness of care across all --

across the entire hospitalization.

MS. JOHNSON: Any comments on that from the developer?

MS. WATT: Wouldn't it be easier if we moved back over there? Okay, I'll do that. Are you asking us -- I'm sorry, what was the question?

MS. JOHNSON: One of the members has suggested that you actually build a composite measure that would include both of these things. Or an actual all-or-none, yes. I'm sorry.

MS. WATT: Well, sure, we could make one measure out of the two of these. It doesn't change anything in terms of the burden of data abstraction or anything else. They are already completely harmonized in terms of data elements and data element definitions and that kind of a thing.

DR. PACE: I think the suggestion though is not so much to decrease the burden about data collection or harmonization, but

it's telling you more than each one singly. It's telling you whether each patient got both things that are appropriate. Is that what you're getting at?

MEMBER KAPLITT: The question is, and this is the question that you asked is do we have data on -- what percent that get the two-day aspirin don't wind up getting discharged on -- or whatever it is for reasons other than medically appropriate reasons, right? Someone gets an antithrombotic and then they bleed and then they get discharged without it. That's appropriate, but that's going to wind up being excluded anyway or adjusted or whatever.

So what percentage of patients do people start them on antithrombotic and then forget to discharge them on it, or choose not to for inappropriate reasons or something? Because if the data indicated that that was a vanishingly small number then there would be, you know, then it is an unnecessary burden,

the second one.

MEMBER KAPINOS: May I ask -- are we expecting an answer?

MS. JOHNSON: Yes, go ahead.

MEMBER KAPINOS: To me I thought we said that the second one -- so upon discharge the antithrombotic, somebody convinced me when we voted that it was a little bit tied to the correct antithrombotic for the etiological work-up. So upon discharge if you are on aspirin when actually on day three you were found to have a-fib then that's wrong, right? So the -- so if we combine the two then we will lose the fact that the second measure, measuring a discharge was also trying to make sure that the accurate antithrombotic is matching the etiological diagnosis.

MEMBER KAPLITT: So where does it do that?

CO-CHAIR TIRSCHWELL: I guess I'm not sure that it does that. I mean, there is

the second measure about a-fib, but other than that there's nothing specific about the antithrombotic agents. Do you guys want to respond to that?

MS. KOLBUSZ: All I was going to say in that regard is that we do use that table of medications as the doctor pointed out. And for stroke 2 generally it's aspirin that they're receiving on arrival. But there are other medications considered usually at discharge. So you might lose that granularity.

MS. JOHNSON: Jack?

MEMBER SCARIANO: Yes, I think that actually we should have the two of them. You know, I just don't think that actually stroke patients get discharged and they aren't on aspirin. I mean, the overall data that we talked about yesterday, it's just bad data. I just know people in private practice and it isn't showing up in the chart audits because, one, we either just tell them that at home you

should take aspiring or two, is that they're already on aspirin. But as he's saying that actually doctors have discharged patients who have had strokes and they aren't taking aspirin I think is not right. I just don't think that that data is actually valid.

MEMBER WADDY: Actually, if I remember correctly, and I'll have to pull up the information but I think that they did find that in regards -- that African-Americans were less likely to be taking aspirin after discharge. And that's actually why we have a couple of initiatives or research projects right now in order to improve that transfer from the hospital to the outpatient setting in order to improve compliance with that.

Whether or not -- how much of that is due to prescribing habits versus either filling prescriptions or paying for it is not entirely clear, but I don't think the evidence is there to refute that at this point.

MS. JOHNSON: Ramon?

MEMBER R. BAUTISTA: The idea of harmonizing stroke 2 and stroke 5 will have to take into account along with the plans for 0325 as well. We just can't harmonize without taking 0325.

MS. JOHNSON: Michael?

MEMBER KAPLITT: I guess just a -- you know, in isolation yesterday part of what we did with 1b was we looked at, you know, what the performance gap was. And rereading this, so the performance gap for 0435 you said was something like 2 percent, and part of our discussion was those are small percentages but big numbers of people. So it was about 2 percent for 0435 and it was 3 to 4 percent for 0438.

So the question is is it largely the same 2 to 3 percent. That's the fundamental issue because in isolation we looked at it that way and we felt that there was a need. But now the whole purpose of this process is to say are those the same people.

Do we really need to spend our time? So the question is is there data on it, could we generate. I mean, that's I guess what we'd have to ask the developer.

MS. JOHNSON: Would the Joint Commission care to speak to that point?

MS. WATT: Are we still discussing 0435 and 0438? Because we just sort of threw in 0325 too. But if we're talking about 0435 and 0438 yes, we can look at the data to see if patients meeting one meet the other.

MS. JOHNSON: Okay, thank you. So I think what I'm hearing is that the committee has said that creating a composite measure might be something that you'd be interested in seeing. And it sounds like the Joint Commission could actually give some information about that.

CO-CHAIR TIRSCHWELL: Just one final comment. I realize that the data burden isn't any different if you combine the two and you either meet both or you don't. But it is

true that that is a higher standard. And so almost by definition even if there's only 90 percent overlap there's going to be more fallouts on the new measure than with either one combined.

And we're not supposed to probably dictate too much, but I mean, it's a way to sort of take the ceiling down a little bit and leave more room for improvement in an area that we all thought was extremely important but was getting towards the top. So now you could move it down and leave a little bit more room for improvement.

MS. WATT: I think it's a point well made.

MS. JOHNSON: Thank you. I think we're going to go ahead and go to the next set of measures.

MEMBER SCARIANO: I've still got one more comment. Yes, I still think that we should have both of them. Again, even the NIH data, it's probably due to actually lack of

access. But actually my point is that doctors in private practice or neurologists in private practice, and even internal medicine doctors, we do not send your stroke patients home and actually not give them aspirin. It's shown in the charts that actually maybe it's not in the charts. It is higher at actually two days is because everyone writes it in the charts at actually two days.

But if they already have aspirin at home or we just tell them, I know I just tell my patients just take, you know, 80 mg of aspirin. And I usually don't write it down. The actual discharge data is not accurate. I just think that patients who have stroke, they always go home, I'd say 99 percent of them go home on aspirin.

MS. JOHNSON: Okay, anymore comments before we close this one and go to the next one? Okay, let's talk about the measure 0325 versus 0435. So to remind you one is an AMA-PCPI measure measured at the

clinician level, the other is the Joint Commission measure, the discharge on antithrombotic therapy that we were just discussing.

So the differences besides the level of analysis obviously is -- probably one of the big differences is the TIA group in the population of the AMA-PCPI measure. And also, obviously I'm not a clinician. I couldn't tell if it was the same list of drugs or not.

So, I guess the first question to ask is can one be considered superior or do you feel the need to have two different measures. Bill?

MEMBER BARSAN: Well you know, another difference was that the one also includes TIA patients and the other one doesn't.

MS. JOHNSON: Right.

MEMBER BARSAN: So that's kind of a big difference.

MS. JOHNSON: I'm sorry, I thought

I said that. But you're right, that is a huge difference. Right. Ramon?

MEMBER R. BAUTISTA: Again, the fact that you have two different levels here is like comparing apples and oranges now. I mean, I don't think you can really compare and say, you know, one is better. They are two different measures as far as I can make out.

DR. PACE: I think the question is -- I think the first question is is this -- which you've already answered by saying the measure was suitable for endorsement is that it's appropriate to measure at both the clinician level and the facility level. So I think by moving those measures forward that decision has already been made.

So there are differences and the question is -- and I think we'll just grant it here because they come from different data sources that in reality they can't be combined into one measure in the current environment that we have.

So the question is are these differences indicated? Should one include TIA and the other one not? Should they both include TIA? And are the med lists appropriate? So I think for these we're looking at the harmonization and what does the evidence say really should be the measure focus and denominator population.

MS. JOHNSON: David?

CO-CHAIR TIRSCHWELL: So I would suggest that the 0325, the PCPI measure, give strong consideration to removing the TIA. I think although I'm not suggesting TIA is not an important condition I think there is so much as we call it "squishiness" in the diagnosis of TIA that you really could be led astray by that. And with ischemic stroke it's much clearer and I think is probably a better consistent performance measure.

MS. JOHNSON: Fred?

MEMBER TOLIN: Actually, David, I agree with that 100 percent and for that

reason I would really view looking at stroke and the Joint Commission measure as really in some ways superior to the muddiness that we see with the TIA sort of garbage list if you will.

MS. JOHNSON: Michael.

MEMBER KAPLITT: Here's the process concern I have. So I agree with you in principle. The process concern I have is that if we're agreeing that these are two separate measures and that we're not going to be using this issue to decide on superiority of one over the other because we've already decided we're going to keep both then the question I have is it seems to me we're essentially revisiting something we did yesterday.

If TIA is that much of a problem why did we approve it yesterday? Because we're not now evaluating TIA saying that that's the crux of how we're going to decide between these two. We're saying that we don't

think it's a good measure and that's different.

DR. BURSTIN: I think -- and it's an excellent point. I think part of what's different today than yesterday is yesterday you're looking at a measure in isolation. So today you're looking at it and you're saying will you get valuable results if you have a clinician measure and a hospital measure that are in fact different because the populations they serve are different. So the other possibility would be to ask PCPI to have a stratified rate, one for stroke, one for TIA.

MEMBER KAPLITT: I'm not disagreeing with that, except the discussion we were just having is how valuable TIA is and that, you know, I mean that to me seems fundamental to whether or not that was an appropriate measure to endorse. I agree with your point but that's not what we were talking about just now.

CO-CHAIR TIRSCHWELL: My opinion

is that yesterday it wasn't important enough to not endorse, and that today we're potentially looking at just harmonization and potentially a small improvement. I mean, I think they easily could have taken away from yesterday's conversation that maybe they want to reconsider the TIA thing and the suggestion that giving a stratified rate might be the first step towards convincing them whether or not TIA is a relevant thing to continue to include.

MS. JOHNSON: Greg, is your hand still up? No. I think Jocelyn.

MEMBER J. BAUTISTA: I think if we remove TIA then they're essentially the same metric. You know, so there are minor exclusion, you know, length of stay exclusion, things like that, but then we're essentially the same metric. We're measuring the same processes in the same patients, whether they're discharged on antithrombotic therapy. And the only real difference then is this one

variable of physician, right? So if you just add that one variable physician into say the Joint Commission metric you have all the data you need to stratify it by physician. Why do we need a whole `nother data set?

MS. JOHNSON: Would anybody like to comment on that?

MEMBER BARRETT: I think it was said before that it may be that many hospitals may not report the Joint Commission metric if they're not trying to achieve certification.

DR. PACE: So I think one of the -- that's an excellent question of why you can't have one measure that you can compute performance at the facility and the clinician level. So the question would be whether the, for example, the facility data captures the clinician so that they could actually do that, or whether the clinician-level measure actually captures hospital data.

MEMBER J. BAUTISTA: So we're not advocating who collects the data, right?

We're just advocating the metric and the data elements.

DR. PACE: Well, the problem is that a lot of the detailed specifications are very much tied to a data source. So, ideally you would be able to, you know, if you don't get to that level of detail then you're not exactly sure how exact the measures are. You get even more error in these measures than, you know, just what you normally have.

So that's the reality that we're in is that we have these different data sources and measure developers that specialize in a data source, and we -- maybe when we get to electronic health record measures that will be an easier lift in terms of having measures that can accommodate both. But, I think it's a question that comes up over and over and we might want to have the developers respond to whether their measure could accommodate the other level.

MS. WATT: This is Ann and

presently I don't believe hospitals have the capability of collecting data from physician billing records which is the data source for the 0325 measure. I agree --

DR. PACE: Right, and I don't think that would be the issue. It would be the measure as you specified but having a physician indicator so that then you could compute that measure for a physician level.

MS. WATT: Well, you know, I can't speak to how the data are collected for the PCPI measures. You know, I suppose that that's a possibility. I can tell you that to the extent possible we believe that the data elements and so forth are harmonized between these two measures.

MS. JOHNSON: Dan, you've been waiting for quite awhile.

MEMBER LABOVITZ: I beat this horse yesterday. I'm going to do it again but I'll keep it brief.

I agree with David Tirschwell, TIA

is a quagmire, a cesspool. It's full of silliness. But fundamentally transient ischemic attack and ischemic stroke are the same disease separated only by luck. In TIA you have rapid re-vascularization or excellent collaterals, but aside from that they're the same disease, the same pathophysiology, the same approach to secondary prevention.

Abandoning it because doctors are undisciplined about applying the diagnosis is sending the wrong message and it's not logical. The demand on us is really better diagnosis and I think if we -- I think we have the capacity to look for that from this perspective. I applaud the AMA for saying yes, we've got to do it even though it's bad.

MS. JOHNSON: So I think what I'm hearing right now in the table is in terms of harmonization there's one idea of taking out the TIA patients. The other idea is maybe seeing if the AMA group could stratify so that you could keep your garbage diagnosis but be

able to compare those measures. Is there -- did I miss something? Is something else on the table?

MEMBER KAPLITT: I still think that Jocelyn's point is well taken. Because I think that we're starting to buy into the idea that, you know, inherently there's, you know, the physician measure and the hospital measure are two separate things. I still don't necessarily buy into that point in this case.

So in a situation where your measuring let's say use of antithrombotics in hospital versus use in general practice then there is a big distinction between a physician measure because you're not going to capture all of that with hospital data.

But in this case the definition of both of these are at discharge. By definition discharge is from a hospital. So I still do not understand the distinction between in this particular case with these hospital-based

measures -- I know they're different data sets that we're collecting from it. I don't understand the distinction in terms of what we're measuring because the ultimate goal -- yes.

So in one case you're going to be able to tell individual physicians how they're doing. In the other case you're going to be able to tell the hospitals. But if the goal is to get the patient to -- is to get more patients to get the right care at the end of the hospitalization, then if the hospital data is collected and that hospital is doing poorly they can then get granular on why they're doing poorly.

CO-CHAIR TIRSCHWELL: The only other thing I was going to suggest was I guess I don't know how what we do or what we say can influence things, but you know the idea would be that you wouldn't have two separate, totally distinct pathways for reporting for what is essentially the same data.

And so you know, how do we support the coming together so that there's just one data collection that can feed multiple systems. And I mean, if the suggestion is well, we have to toss one of the measures well then maybe we should. But it seems to me that the goal has to be that this vast array of parallel redundant data systems need to be combined.

MS. JOSEPH: This is Diedra at the AMA-PCPI. Can I comment on any of the discussion?

MS. JOHNSON: Yes.

MS. JOSEPH: So, thank you for the opportunity. I'm sorry I didn't jump in earlier. I didn't know if I needed to wait until the end.

So with regards to inclusion of TIA I just wanted to explain kind of the clinical expert panel's thoughts behind including it in the measure. We did discuss harmonization with the Joint Commission

measure as we were developing the measure and during the maintenance and review of the measure.

The intention was to have a broader applicability and I think as someone stated earlier that the TIA patient -- evidence does support the use of antithrombotics in TIA patients. And the -- even though coders might suggest a diagnosis the physician ultimately has to sign off on it and is responsible for the accuracy of the coding and how the diagnosis is coded.

And so since evidence supports the use of antithrombotics in TIA patients we thought -- and because there is still an existing gap in care then the clinical expert panel decided to leave it in. And again, because of the broader applicability issue.

With regards to an additional question or statement that was made regarding combining the measures or just having one measure versus having a facility-level and a

clinician-level measure, our measure specified at the clinician-level but our measure results can be aggregated at a higher level of measurement.

Still I would -- we would advocate for both measures being endorsed only because it is important to, number one, to capture the information at -- so that clinicians can know how they're doing with regards to accountability and so that hospitals can know how they're doing with regards to accountability. And also because -- sorry about that. I'm losing my place. Also because the measures are included in different national programs, and things like PQRS and meaningful use. And so getting rid of, or losing the endorsement for one of the measures might cause the measures to no longer be included in national programs. And it's obviously important for the data to be collected in order to improve quality. So that's what I had to say about that.

I don't know if anyone had any specific questions for me but now that you know I'm on the phone feel free.

MS. JOHNSON: Jocelyn?

MEMBER J. BAUTISTA: So I completely agree with what Dave said earlier. I would vote for getting rid of one of the measures and I would vote for keeping the Joint Commission measure.

MS. JOHNSON: Mary?

MEMBER VAN DE KAMP: My concern with getting rid of the -- of just keeping the one Joint Commission is you haven't gotten to the physician level. And so are you saying add to?

MEMBER J. BAUTISTA: Right. So you know, when my hospital sees that one physician is not performing the way it's expected they let that physician know. So that data would not be lost.

MEMBER KAPLITT: And in reverse if a hospital is at 100 percent compliance then

what more do you need to know.

MEMBER VAN DE KAMP: So you're going to drill down from the facility.

MEMBER J. BAUTISTA: Hospitals will, yes.

MEMBER KAPLITT: If a hospital is being told that they're lousy and that's affecting their thing, they're going to get into it, believe me. Whereas vice versa, individual physicians may say all right, I have special reasons why I'm, you know, different.

MS. JOHNSON: Helen?

MEMBER COONEY: And you could actually look at the physician level in the outpatient setting once they're really established back in the community rather than at discharge and perhaps get complementary information.

MEMBER BARRETT: I just want to remind everybody the same thing I said before, that the performance gap is really different

on these two measures. So we really don't know that all the providers were captured in the facilities.

DR. BURSTIN: Just to add in, I think this is a great discussion. It sounds like it is a future tense as opposed to at the moment. So the question I would also have is can we ask the Joint Commission and PCPI to think about if there are some opportunities to potentially have clinician-level indicators out of the Joint Commission measure which seems ideal.

But also do keep in mind one of the reasons for the PCPI measure which they have harmonized to the extent of at least with the exception of the TIA issue is that it allows physicians to report through PQRS. So it is at least harmonized and gives them, particularly neurologists I think a measure to put forward as part of that program.

If it is fully harmonized and it's information that's complementary it seems more

to me at least like a longer term issue to potentially ask the Joint Commission and PCPI to bring those together so you could actually extract from what one of them does information on both.

Again, the optimal situation is to be able to cascade up and down to understand where there are issues and where there are problems and point towards where improvement needs to be.

MS. JOHNSON: Jolynn?

MEMBER SUKO: I guess one thing that I would say is that the administrative burden of collecting this on both sides and the difference in validity as has been pointed out I feel like I'm again beating a dead horse, so probably annoying.

But one thing I would suggest to the AMA-PCPI is that these are the same patients that the Joint Commission is abstracting on. And can you test your measures at a patient level against the Joint

Commission's measures? Because my hunch is that many of the same -- if you took one practice, that practice at a hospital who was participating in PQRS and you took the Joint Commission measures you could measure -- that would be an effective test of the validity and the opportunities to reduce that administrative burden.

MS. JOSEPH: I'm sorry, this is Diedra at the AMA. Could you further clarify what you're saying? I got a little confused because you said hospitals participating in the PQRS system but it's actually at the individual physician level that data is submitted to PQRS.

MEMBER SUKO: Thanks, I -- what I'm saying is that you're measuring conceptually the same thing. And likely those patients that you're submitting to the PQRS, while you're doing it on a different -- with a different method and claims form, the hospital where that care was delivered is also

submitting data on that same patient. And so a test would be to cross-reference what's submitted through the physician claim form and the hospital on that patient level.

MS. JOSEPH: What you're saying makes sense now and it's clear what you're asking, but that's something that I can look into with our testing team. But I'm not sure that that can happen in the near immediate future. But I can definitely pass that suggestion along.

MEMBER SUKO: Okay.

MS. JOHNSON: So Karen and Helen, a process question. Where do we go now?

DR. BURSTIN: I want to hear Dave's comment first. Then we'll --

MS. JOHNSON: Sorry.

CO-CHAIR KNOWLTON: I'm not the clinician here but this strikes me -- I have a process question. This strikes me as compared to the robustness of what we considered in the past day and a half this

appears like NQF light to me. You know, it just strikes me that we are taking an action that can result in some drastic changes to these measures with nowhere near the depth of consideration that we heard in debate for the past day and a half. And I'm uncomfortable with it.

I don't know a lot of the things you're talking about, you're more expert at it than I am, but from where I sit I'm trying to figure this out. It's a little dizzying. Not all the developers have time. They're defending their measures. And I'm saying shouldn't we have the robustness to this process that we had to the other process. I mean, maybe it's a little more work but shouldn't a group look at this and dig in?

I mean, I learned a lot in the past day and a half. I'm not learning now and sort of people are winging it. And I don't mean that offensively but people -- my gut reaction, very different level of discipline.

And it troubles me because what happens as a result of that, we say okay well this one's gone and this one will stay and we'll split them. I feel like I'm playing Let's Make a Deal, you know.

(Laughter)

CO-CHAIR KNOWLTON: That I know how to do. So I'm troubled with the process.

DR. BURSTIN: I think that's very fair. And you know, the whole concept of doing related and competing is still relatively new, really the last year. I think we were really hearing from the field please stop the cacophony of measures at different levels that aren't harmonized. It drives people insane when they get measures from different health plans and hospitals that tell you your performance is different. So we've really taken this on. But I think those are fair concerns, Dave.

And actually what I was going to suggest earlier is I think in some ways the

developers have now heard this discussion. I think at this point it should be up to them to talk and bring back a response to the committee that you can read and review in a time period that's not quite as rushed as doing it at this moment.

MEMBER KAPINOS: I would have loved to see them at the front end meet up and say hey, we have this common measure here, let's get together and talk about this first before the committee and then we can decide if it's okay or not rather than at the end trying to combine two not necessarily the same measures and try to figure things out.

DR. BURSTIN: You're speaking our language, yes. We completely agree and in fact some of you may have heard that NQF is about to pilot a new two-stage endorsement process that would bring measure concepts in first and allow for that harmonization. And then a fully spec'd out, tested measure to follow to try to avoid some of this cacophony.

MEMBER WADDY: I just wanted to add that I agree with that. I mean, we don't know the impact of these minor and in some cases major suggestion changes and how that will impact the data that we would see on the final end.

MS. JOHNSON: Okay, this has been a great discussion. So I think if no other words on this group let's go onto the next group.

And I'm keeping an eye on the clock. I think in some ways the next group of measures will be similar so we night not have to rehash all of the discussion but some of it we definitely will. So if you go to page I guess 5.

DR. PACE: So one of the things that we might do, and if you know are these kind of the same issues? Between the same developers? What page is it?

DR. BURSTIN: Page 5. Isn't the next measure discharged on? No.

DR. PACE: VTE prophylaxis.

MS. JOHNSON: I learned yesterday that you're supposed to say VTE, right? Yes. So that is -- I think I am -- I am on page 4 of my sheet so I may have a little bit different sheet than you do.

Okay. That new document is the actual detailed specs which is a little different than what we looked at yesterday.

Okay, so the detailed specs, let's go ahead and start with 0434 and 0371. It is a little different in that 0371 you guys did not look at. That was not one of the measures that you looked at in your project. We did not look at 0371 yesterday when we discussed VTE measures. So, what 0371 is is another measure put forward by Joint Commission and it's not exactly related or competing, but it is -- I'm trying to figure out.

DR. BURSTIN: Karen, we could try to make this a little bit easier since we know all these measures. So the biggest

distinction here is that, and again the Joint Commission can help here, but the measure 0434, the one we talked about was VTE prophylaxis in the setting of stroke. And then there's a broader measure. And my understanding, correct me if I'm wrong, is that that broader measure excludes stroke. Is that correct, Ann?

MS. WATT: I apologize. I don't know the NQF numbers. If you could give me the name of the measure that would help. The VTE prophylaxis from the VTE measure set. Could the two be combined, that's the question?

CO-CHAIR TIRSCHWELL: I was looking through it and other than that one excludes stroke and one is only stroke I was really hard pressed to find much of a difference.

DR. BURSTIN: The PCPI measure I believe is related to surgery.

MS. JOHNSON: It's not only

surgery, it's a broader measure.

MS. WATT: Sorry I didn't bring my sign over too. Even though -- you're right, the measures are very similar. The big difference between the two of them is that two holes is okay for your general non-stroke population whereas that's not true for stroke. And that's the difference basically. Am I right? Yes?

MS. JOHNSON: Ramon.

MEMBER R. BAUTISTA: I don't mean to be a pain, but I think it's fundamentally unfair to ask us to assess a measure that we looked through yesterday in great detail, in fact to compare to something else that we saw yesterday. So I think it's actually unfair and probably invalid exercise to do this.

MS. JOHNSON: Okay, any other comments? Okay, nobody wants to comment further on that. Let's go onto 0240 versus 0239 which are related measures. And maybe you would have the same question on this one

because 0239 you did not look at in this project.

DR. BURSTIN: I actually don't know that they're really that related. I think again it's the same distinction we just talked about. One is for stroke patients and one is for the general population. I think it's the same issue. I don't know that we need to do much more on that at this point.

MS. JOHNSON: All right, well then we're going pretty fast here. Then we'll go to 0240 which is the AMA-PCPI measure and comparing that to 0434. Both of those you did look at yesterday in this project. They are competing measures but they are also pretty much it's the same thing that we had earlier which is one is clinician, one is facility. So the third row on the big screen up there, 0240 versus 0434.

MEMBER KAPLITT: Well, it's the same as the other thing. One is physician and one is hospital. It's the same as what we

just did for venous thromboembolism.

DR. PACE: We can just ask the same question to the developers on that to come back to you with that.

MS. JOHNSON: Okay.

MS. WATT: Could I just -- I'm sorry -- make the distinction between the two measures? The 0240, it looks at DVTs only whereas the 0434 looks at DVTs and thromboembolism.

DR. BURSTIN: But the prophylaxis is the same.

MS. WATT: Okay.

MS. JOHNSON: So I guess the developers, we would just ask you to maybe make a response to the committee about whether you think there's a possibility that you could aggregate them differently so that we wouldn't need to -- just give us your opinions on that. Just like we did with the last set. This would be at a later time, not today.

MEMBER KAPINOS: If all the

evidence that they look at is actually from trials that looked at VTE I think it's just that they did use the wrong term by saying VTE prophylaxis. As I said yesterday the majority of the trials is calling it VTE because what you're really preventing is DVTs and/or PEs.

So if actually all the evidence in those trials -- in those two measures are the same then actually it's very simple for the Joint Commission. They should remove every -- in all their documents wherever they're saying DVT just change it to VTE and you're fine.

MS. JOHNSON: So one extra suggestion, that AMA-PCPI may want to consider renaming their measure. Okay. Any other discussion on this measure group?

Okay, let's go onto measure group 3, anticoagulant therapy. What we have here is a group of three measures, and the first one 0241 versus 1525. Again, with 1525 that is a measure that you did not consider in this project. It is a measure that is broader I

think than the 0241. So the question there is is it necessary to have a separate measure for just the stroke population. And again, perhaps you guys don't feel comfortable making any response on that.

MEMBER KAPLITT: I would argue these are -- I mean one is a hospital-based. This is what I was referring to earlier as being totally different. One's a hospital-based and the other one is in the ambulatory clinician's office. So we can start getting into all the little fine details, but it's so fundamentally different that, you know, if we didn't review it I don't see how it's even relevant to try to do that.

MS. JOHNSON: So you feel that they're so different that the first question, both are needed and they probably couldn't be combined.

MEMBER KAPLITT: Well sure, because one is looking at how patients are, you know, what we reviewed or what's happening

to patients who are being discharged, how they're being handled upon discharge. The other one is about general office practice which has nothing to do with that.

MS. JOHNSON: Okay. Any other comments?

Okay, let's go to the next set, 0241 versus 0436. These two you did look at yesterday and they are for the most part the same discussion as we've already had. One is a clinician-level, one is facility-level, both are at discharge. And I think other than the setting the other difference I think might be the definition of having flutter in the numerator. I think that was one difference in terms of potential harmonization. NTIA, thank you. David.

CO-CHAIR TIRSCHWELL: I would just comment that again I'd urge the -- one is Joint Commission, right, and one is AMA, to consider harmonization as best possible with these slight changes in definition. And also

whether in some brave new world in the future some of the data collection burden could be harmonized so that it only has to be done once through one method, and again be easily divided out to serve both masters.

MS. JOHNSON: Any other comments? Okay.

MS. JOSEPH: This is Diedra at the AMA. Can I provide a comment about atrial flutter?

MS. JOHNSON: Yes. Sure.

MS. JOSEPH: Thanks. So the reason why we opted not to include atrial flutter, because we did consider it during the clinical expert panel discussion, is because we use the updated guideline. As you know, our measures are based on guideline recommendations, and we use the updated guideline published in the Journal of the American College of Cardiology. And -- to support the measure.

And there was a 1c recommendation

referring to atrial flutter. And that recommendation actually was based on expert consensus only and so the clinical expert panel thought that they should only focus on the patient population with the strongest evidence supporting it. So that's why we limited the measure to atrial fibrillation only.

MS. JOHNSON: Thank you. Do we have any other comments about this set of measures?

Okay, let's go to our last measure group. And this one is the rehab services ordered and assessed from AMA-PCPI and Joint Commission. And again, it's pretty much the same question. One is facility-level, one is clinician-level. And I'm not sure that there are major differences in the definitions here.

MEMBER KAPLITT: I mean, they're slightly different, right? Because one of these is the order -- is that it was actually ordered. The other is that the patient was

assessed which are not totally equivalent. And there may be value in both. I mean you want to make sure that they're getting ordered, but you also want to make sure that they're getting evaluated. So whether they both have to be captured, I don't know, but you know, I assume if people are getting orders for it without being evaluated that's probably not a good thing. If people are being evaluated and then it doesn't go on to an order that may be okay. So you know, it may be the evaluation is more important, I don't know.

MS. JOHNSON: So these are related measures. Is there any issues of harmonization that you guys want to bring forward to ask the developer to respond? Okay, sounds like no. All right, good. So we've gotten through a very difficult session.

So I think the last order of business today that we need to talk about is measure gaps. So I know especially yesterday

but even somewhat today there was some discussion from many of you on ideas for other measures that perhaps could be considered by developers. So we wanted to open up this time so that if there's any other ideas that you may have that would be great things to measure for quality in the stroke field please bring those to the table now. David.

MEMBER HACKNEY: Well, we eliminated the only imaging quality measure we had yesterday and I just would hope that we can get back to that topic. I think there's a lot of important things that could be done. It was more the details of what that one included than it was the principle that the acute imaging is important.

MS. JOHNSON: Great. Did you have any like --

MEMBER HACKNEY: Suggestions of what it should contain?

MS. JOHNSON: Suggestions, yes.

MEMBER HACKNEY: I mean, it would

get a lot into the details. It might be better to pass on. But I think from the discussion it has to start with something that actually would have an impact on patient care. So it would be how fast was the imaging done, how fast was a reliable interpretation delivered. The time window would have to be appropriate to the time window that's relevant for acute stroke patients, not out to 24 hours.

It might require capturing things like revisions to a preliminary report. It may or may not want to delve into CT versus MR because it's such a controversial issue, but it could at least provide some guidelines about a minimum imaging study that should be done in the acute case. But I think all of those would be potentially of serious impact to patient care and we ought to ask that to come back with something that's more refined in form about our discussion yesterday.

MS. JOHNSON: Thank you for that.

Dan.

MEMBER LABOVITZ: This may be a question -- well, I'm not sure who would address this. But we're confronted I think with, what we've seen over and over again in the past day and a half is very different data sets some of which are generated by chart review in the hospital and based on what the doctors wrote down interpreted by rules that are sometimes radically different between what doctors think and what coders think.

A doctor can write down "Blood cultures positive times 4 bottles" and everybody knows what that means, that patient is septic. And the coder will say "UTI" because their rules are different.

MEMBER KAPINOS: -- sepsis is bacteremia.

(Laughter)

MEMBER LABOVITZ: We face here a problem of -- in the end we generate ICD-9 codes and submit data for billing. We submit

data for quality review. We submit data for state review. I would suggest that what's never -- what never seems to happen, and there's no attempt to do it is a systematic approach to confirming that what gets reported in say Get With the Guidelines actually matches what gets submitted say to CMS with an ICD-9 code for billing. I think they can be radically different.

And what I see in my community, my area, there are hospitals that turn in enormous amounts of billing data and not a whole lot of quality data, but they get gold stars for all their quality. There's a real mismatch there and what I'd love to see is a measure -- some attempt to measure quality reporting to voluntary -- voluntary and quality reporting to say Get With the Guidelines and actual hard data that gets billed for.

MS. JOHNSON: Okay, thank you. David?

CO-CHAIR TIRSCHWELL: And this partly may demonstrate my ignorance about other quality measures in other areas, but end of life care in stroke is tremendously important. And practicing at a tertiary care place where a lot of our stroke patients die, there's I would say even amongst ourselves there's a tremendous amount of variability in the quantity and quality of care that goes into the dying process.

And I don't know, maybe you know, Gail, if measures exist, but I would like to see that sort of play out into the stroke arena more specifically if possible.

MEMBER COONEY: There are palliative care measures but I do not believe that there are any that are stroke-specific and that's what I was up for, was to suggest that we develop, you know, given the stroke severity rating the presence or absence of a palliative care consultation. Because currently 85 percent of hospitals have

palliative care available in them and that would be one way to approach those end of life issues.

MS. JOHNSON: All right. Jane?

MEMBER SULLIVAN: In part in response to the -- what happened with the speech-language pathology measures and in part from where I sit as a physical therapist I'm really interested in functional outcome measures, both positive and adverse like falls data, those kinds of things. So I'm not sure where that is in regards to this process, but I'd really like to see some attention especially looking at stroke severity relative to functional outcome.

DR. BURSTIN: -- just mention that we're doing a meeting at the end of July actually that Karen's leading focused on the methodologic issues and looking at delta as function, for example, and other patient-reported outcomes. There's a lot more that needs to happen there but there's a lot of

methods that really need to get cleared up first.

MS. JOHNSON: Gregory?

MEMBER KAPINOS: I want to support the idea of Dr. Labovitz. It was really enlightening to come here and see that actually all the data that were based on these -- from billing and coding which is sometimes not perfectly accurate.

The other idea that I think I shared with Dr. Tirschwell yesterday was once those measures are in effect and we collect just a rate or 90 percent of our patients are getting, for instance, antithrombotics upon discharge I think that this absolute value should be actually weighted with the fact that some hospitals will exclude a lot of their patients.

So as we were discussing yesterday, when you read for instance a randomized clinical trial you have a flow chart that tells you how many patients were

excluded, then how many patients met the inclusion criteria, and then they were randomized into the two arms. For your absolute number -- so I would -- what I want to get to is hospital A is not necessarily better than hospital B because their score is 92 percent versus 90 percent, if actually hospital A excluded 90 percent of their ER visits for stroke for instance. So maybe we should actually look at the number of exclusion of patients as a way to weighing the score. So that 92 multiplied by 50 percent of exclusion comes up to a new score that is actually more valid to compare to hospitals.

Because otherwise, as I said, I think a lot of hospitals could have a tendency to exclude a lot of their patients so that 90 percent compliance to antithrombotic for instance is actually generated on a very small percentage of their actual patients. So I think NQF should revisit the issue of that absolute value and maybe create a more complex

system looking at the exclusion.

CO-CHAIR TIRSCHWELL: Can I respond to that just briefly? Because I slightly objected to that suggestion yesterday because I think that different hospitals will have different appropriate exclusion levels. And so the hospital with the more complicated patients and the more appropriate exclusions in that system would be counted against.

And you know, it costs a lot of money and I suggested that Greg start a company to do this, but validating the exclusions, external validation of the exclusions would be an extremely interesting point. And I think there would be some shocking discoveries potentially, although hopefully in only a small number of hospitals. And I do have a couple other points but I think --

MS. JOHNSON: I think Anna was next in line.

MEMBER BARRETT: Thanks. Nobody's

going to be surprised to hear me say that we shot down our only rehabilitation measures so both outcome and process measures in rehabilitation given the amount of funds.

I started my career in the VA, or I was trained in the VA and they at least always say, you know, rehabilitation gets this much of our budget because that's so much of the person's life and expenses, and here's the amount that we put into research and all the acute care stuff. Well, that can take care of itself.

Lastly, I would also say that hidden health disparities play a large role both at the quality of care -- quality care we hope identifies hidden disabilities after stroke or non-motor disabilities. After Parkinson's disease non-motor became such a buzzword that we can talk about it in stroke as well.

But as we're talking about a lot of measures one has to have communication

ability, one has to have appropriate ability to interact with the physician in order to participate even in the measure, and then of course falls, medication adherence, and a number of safety measures are determined by hidden disabilities.

MS. JOHNSON: Michael.

MEMBER KAPLITT: I would just ask NQF for the next phase or round or whatever if we could in advance have some sort of a summary of other measures that we haven't evaluated that are already endorsed in this area.

For example, the radiology question, yesterday we were told that there actually is a measure in terms of the speed with which a radiologist reads the film or something like that. So having a summary of all those, not the whole extensive thing, but just maybe the title, the inclusion/exclusion -- I mean, the numerator, denominator and exclusion criteria, just the fundamentals

summarized in a table in advance would help with this kind of discussion I think.

MS. JOHNSON: Thank you. David?

CO-CHAIR TIRSCHWELL: So, two comments and these are more I think process suggestions for NQF. So, I feel bad about all those speech pathology measures that went down. And I think that a pre-review process potentially with NQF staff to identify a weakness in our application without actually changing any of the truth behind it all could have prevented that. And so I don't know that you have the manpower to do this but I think that it would make sense.

And quite honestly, you know, there were all these updates to the forms after the conference calls, but maybe the first set of updates should be before you show the applications to us so that they're in better shape when they get to us. There were some real weaknesses in some of the original applications that I looked at too, so

something for you to consider.

And then as far as the competing and harmonizing related measures I think as you guys probably noted today, I think the way we have this really structured approach to identifying and voting, I think you need to create a little more process around that discussion to help organize the way things go.

MS. JOHNSON: Did Helen or Karen have any response to those before we go on?

DR. PACE: Yes, just, you know, the pre-review is something we've identified and we're actually piloting some processes to do that. Because it is a consistent issue of the quality of the application itself. And you know, right now our time lines don't allow that which has been a problem so it kind of repeats and repeats. So we are pilot testing doing some of that. We agree that it would be helpful.

MS. JOHNSON: Bill.

MEMBER BARSAN: This may reflect

some of my ignorance about how the process is carried out because this is the first time I've been involved with NQF, but are you mostly passive in terms of waiting for people to come to you for things, or do you actually go out and solicit new things? I mean, I would suggest that if you don't it would be -- like for example, some of the suggestions we made, will you go out and try to get people, solicit people to turn those in?

DR. BURSTIN: Right. So we try as best we can to go out there, let people know projects are upcoming. Part of what we've tried to do moving forward is actually having a schedule for when projects will come up at a regular basis. So the field is just on notice when they can submit at various points in time.

The hardest thing actually frankly is that there is a limited amount of money out there for measure development. And so I think the challenge is taking some of the great work

emerging out of research and trying to translate some of that into measurement.

But yes, I mean anything you guys could do particularly before this next round of measures to kind of let people know out there that there is an open call for measures, we'd love to get stuff in.

Again, the other thing I pointed out is I think there's still a lot of effort being made around the similar set of measures. And we haven't -- I think we're hoping to see more that kind of takes us to a very different level.

MS. JOHNSON: And a plug for next phase. We do not at this point have any TBI measures in our pipeline nor do we have any migraine or headache measures. So if you guys know of any folks who are working in that area please publicize the call for measures.

MEMBER WADDY: I was wondering is there any data on how much cost gets added by having these types of measures as well as --

so really what's the burden to the hospitals as well as what's the overall impact and how much do they change things?

DR. BURSTIN: It's very variable depending on the kind of measure. Obviously when measures are completely claims-based outside the hospital that actual collection of data is not something that's a burden on the hospital but still is in terms of reacting and kind of improving around it we hope is a significant part of it.

We for awhile there actually were asking developers to let us know how long it took to collect the data, the costs of it. Just, it's so not comparable across measures that we don't go there anymore.

In terms of impact it's the right question to ask and actually we just recently updated our usability criteria that'll go forward in the fall I guess which is much more explicit about use and usefulness. So, what is the use for the measure and what is the

evidence that it has either improved care or evidence of unintended consequences. So I think we've tried to make that more crisp because I think committees have told us at least to date it's not very crisp.

MEMBER WADDY: I think that would be very interesting, particularly in measures that were pretty close to that threshold such as the antithrombotic use and that 2 to 5 percent, that could weigh into how much of a -- we think an impact can actually be made which hopefully there's a great impact.

DR. PACE: And I think your question of cost also gets at one of the drivers behind the interest in harmonization and competing measures. Because it just adds burden, but the difficulty as you saw is that at this point you all get measures, NQF gets measures that developers have already invested time and resources into. And so it's, you know, we keep trying to move it upstream, like have those discussions before you come to NQF.

To date that hasn't been real successful and we're going to be looking at -- and any suggestions on that would be welcome, but looking at how to continue to push on that.

MS. JOHNSON: We're getting very close to our time. I think Risha definitely had her hand up, and then Jane, and Dave? Oh, Dave, okay. All right. Risha?

MEMBER GIDWANI: Thanks. First off I just want to say I think this is a really thorough and systematic and well-organized process, so thank you, NQF. I am really happy I was able to be a part of this.

I have a couple of suggestions for the next go-around and that's that if there are measures that are outcome-based and that have already been endorsed I would suggest that it be a requirement that the developer show the data since the last few periods of time so that we are able to really assess the impact of these standards. And then the other suggestion that I have is, you know, just to

respectfully suggest that there be a couple of folks with expertise in risk adjustment statistics or economics if there are going to be risk-adjusted outcomes that are going to be evaluated. Risk adjustment is a very sophisticated field, it's also relatively new and it's fraught with a lot of complexities. And I in no means wish to say that we shouldn't be engaging in it, in fact just the opposite, I think we should certainly be trying to help move the field forward. But statistics is a science that has a strong element of an art to it, and so I think that it's just like in any academic or intellectual enterprise, it's worthwhile to have a few different folks with expertise at the table so that we can make sure we're engaging in strong intellectual debate. Thank you.

MS. JOHNSON: Jane?

MEMBER SULLIVAN: Not to beat a dead horse but I want to go back to the speech measures and the work group. And I think

maybe a suggestion. I think during the work group call the group was concerned about the level of evidence. And we were given some guidance to use our clinical judgment as well as the very limited information that was presented by the developer. And that seems like it was a little different than what happened here. And so just further clarification on what the threshold is, what the bar is for the outcome measures. I think that would be very, very helpful in the future.

And I also, I sort of echo your concern about what happens with the suggestions or the queries that the work group make of the developers because I believe that there were some queries made in our work group call that we didn't see addressed. And I think that might have helped and might have resulted in a different outcome at this level.

MS. JOHNSON: Thank you. David.

MEMBER HACKNEY: Echoing what

other people have said, I got the impression a lot of the developers were surprised by the number of questions about evidence of impact. And I'm not sure they were devoting nearly as much attention to that issue as we were. And I think better communication about what we are looking for could have prepared them to present the sort of information that we said was essential without which we weren't moving forward.

MS. JOHNSON: Okay, any other comments?

CO-CHAIR KNOWLTON: Never sit beside someone who's supposed to call on folks. I don't know how to measure this. Again, a lot of it's colored by my own experience. But we have no measure of pre-hospital care. We all agree it's critical. States that have an integrated stroke system and stroke code system have integrated pre-hospital care. I was the beneficiary of that in Connecticut. And it's made an, I believe,

an amazing difference in outcome. And yet we don't -- so I don't know who to measure this.

And you guys, particularly you clinicians are a lot more adept at how you could measure that. But it's such an obvious gap to me that we're not looking at some way to capture that. And I do believe it makes the difference between being functional or being in a SNF unit.

MS. JOHNSON: Michael? Sorry, Salina.

MEMBER WADDY: Or actually, I mean, post-hospital care in terms of being sure not only that the patients receive their antithrombotic prescription or their statin prescription, but whether or not they're actually in control some reasonable amount of time, 3 months, 6 months, whatever you want to choose. And I think that that's extremely important as well. And we do a horrible job at that.

MS. JOHNSON: Bill?

MEMBER BARSAN: Let me just -- hospital care thing. So, pre-hospital care is a quagmire for trying to decide anything because knowing where the responsible people are if you're going to measure quality, who is responsible, it's tremendously variable by state, by city. I mean there's really no -- you have volunteer squads, you have fire department squads, you have, you know, non-profit foundations and all kinds of stuff. And they deliver to so many different hospitals that it's not any one hospital that's responsible, or one group of physicians even that you can say well you're the ones that are supposed to be doing that. And so it's a really difficult area to work with.

I mean, just trying to make sure that all your pre-hospital providers do a pre-hospital stroke scale is very, very difficult. Some of them you can, some of them you can't. But I agree with you, there's the need for that, there's no question.

CO-CHAIR KNOWLTON: If you measure it you can manage it.

MEMBER KAPLITT: Well, the ones you could measure. It's a good point. I mean, you can measure simple things to see how you're doing. Instead of measuring to lay blame you can measure to see how you're doing, right? Like the number of patients that get transferred from one hospital to another, you know.

I mean, you have a setup in various states where you're supposed to do certain things, but that doesn't mean the patients are winding up there. That's true in a lot of areas like TBI and other things. So you could measure the number of patients that would have been candidates let's say for some intervention like t-PA but were delayed because they didn't go to a stroke center, let's say, right? And then that would give you an idea of how you're doing. And then you could measure it by community to see. So you

may not blame anybody, but let's say one area of a state is much worse than other areas. Then they know that they've got a problem with their local whatever, emergency response or something. So, I mean there are some metrics I think that could be incorporated.

I don't think you're ever, you're right, going to get to the level of individual physicians or individual institutions.

CO-CHAIR KNOWLTON: I bring it up because every neurologist or ER doc I've talked to agrees it's a critical factor. Everybody agrees it is a quagmire. In my own experience it happened in Connecticut because Yale took charge and said we're going to set up a stroke code. That's what happened. And they set up stat centers. And my time from a 911 call to my head in a CAT scan was 21 minutes and I was driving on the Connecticut Turnpike at the time of the stroke. And then was transferred by ambulance to Yale 15 miles away and I still had an hour and a half left

on the stroke clock. You know, that was an integrated delivery system and it was incredible.

So, but it was called as a code, it was a stroke code. And so it was run like -- I have a background in EMS. I was a firefighter and I can tell you that when things get called codes they act in different ways because it's a firm protocol. And it just strikes me that something that everybody seems to agree has so much to do with outcome and we don't measure it. So I agree with Michael's addition that we've got to measure it. I also know it's very difficult. I mean it is very difficult.

MEMBER BARSAN: You can measure it, the problem is who do you hold accountable. That's the difficult part.

CO-CHAIR KNOWLTON: Well, maybe at a starting measure it's not, like Michael said, it's not seeking to blame, it's saying how are we doing, you know, and that's the

first step to doing it. But I just notice that it's -- I've never seen it. I raised this at the end of the last stroke session saying that this ought to be --

MEMBER BARSAN: I mean, do you all -- are there any measures at NQF that have anything to do with pre-hospital care? Like with MI or anything else? Pre-hospital EKGs, anything like that?

DR. BURSTIN: No. But we have been doing some work for ASPR, the Assistant Secretary for Public Health Responsiveness -- I always forget exactly what the acronym is -- that does emergency preparedness and actually tried to do an environmental scan for them of what measures are out there around crowding and diversion and some of the sort of systemic issues I think that might lead to some of this. There's very little out there. So we're continuing to see what could be developed in that space.

And again, some of those are

really intended to assess a region, to go to Michael's point, as opposed to a doc or an EMT service, that at least if you start getting data at your region you can kind of, again, drill down to figure out where you can make a difference. But maybe we'll get you on that extra panel.

MEMBER BARSAN: Well, the biggest question is who do you even ask. Who gets the data? Where does the data come from? That's a real fundamental question which will be very difficult.

CO-CHAIR KNOWLTON: Well, that's -- thing about pre-hospital care, especially as it activates the EMS system is there's a lot of data. Because the call comes in, it punches in, it's time-stamped. Arrival is time-stamped. So there's a lot of data. They don't do anything with it but it's a lot of data because all that stuff is legal stuff.

MEMBER WADDY: There's a lot of data for only part of the country largely

because there are huge swaths which is actually a health disparities issue of the country where there is no EMS system and it's completely disorganized.

MS. JOHNSON: Okay, and Bill, your card is down. All right. And Mary.

MEMBER VAN DE KAMP: I had one thing to add. What made me think of that, David, was your comment. And that is that I think it's culturally something we need to embrace more and that's to look at the quality for the improvement of quality rather than the fear of penalty of being the lower half. And I know that's not this group because this is the best of the best, but I think as move forward that's where quality I think is. And I don't know if pay-for-performance ends in coming into that impact but I think it is difficult.

As people look to be measured it's the fear of being measured as poor without the right measurement, rather than trying to look

at what we may measure could improve quality. And so just a comment on, as we go forward, how we sort of think about measuring quality. And the first response is did we do something wrong.

MEMBER WADDY: That's why I think it would be very helpful to have information on how hospital practices or physician practices have actually responded so that, you know, potentially we can even look back through measures, measures that may not have really made a difference in quality but certainly may be burdensome. But there may be other measures and how were those developed so that they actually led to our overall goal, that it's not just a measurement process that we want to go through but an improvement in quality.

MEMBER KAPINOS: I think Dr. Knowlton would say that. But earlier on when you talked about the lack of rigor for this process of this afternoon compared to earlier,

I thought you would have suggested so I'm going to do it now. Why don't we just go through the -- everybody work through the algorithm that you presented on the PowerPoint and you vote for each step? Rather than just put the PowerPoint there and then we all chatted about are we harmonizing or not. So I think we should all work through that algorithm that you put and vote for each step. That would be more rigorous than just a discussion for the harmonization process.

MS. JOHNSON: Okay. When we got you guys going you really came up with several avenues for potential measure development. Is that everything for now? I mean there's always going to be time for you to add things.

Okay, now we need to open up the meeting one last time for any public comments. So Operator, would you open the lines for any public comments? Operator?

OPERATOR: At this time in order to ask a question press * then the number 1 on

your telephone keypad. At this time there are no questions.

MS. JOHNSON: Thank you. So now we're going to turn it over to Suzanne who will tell us our next steps.

MS. THEBERGE: Okay. First of all on behalf of the project team I just want to say thank you so much for all your time the last few weeks. We really appreciate it and this has been an excellent meeting. We've done an enormous amount of work.

So I just wanted to go over the next steps both in our process and for you. NQF staff are going to put together a draft report over the next couple of weeks and we'll send it to you before we post it. But it will go up online we're estimating July 13th for public comment.

And during that time people will have the opportunity to comment on the measures that were submitted, on your decisions, raise any issues that were not

raised, et cetera. That's a 30-day period and following that, that closes in mid-August.

Then we give the developers a chance to respond to the comments on their measures and we'll also give you guys some time to look at the comments that came in. And then we'll have a call at the end of August to discuss all the comments, see -- there may be responses that you need to draft. There may be measures that you need to re-vote on based on new information, et cetera.

And then we go into the voting which we're estimating to start in mid-September. So NQF membership will vote on whether or not to recommend the measures for endorsement. And then the measures following that go to our Consensus Standards Approval Committee and our board for final ratification.

As you know, we will be starting our phase 2 of this project. We'll be sending out a survey in July to see if you're still

available for phase 2. We may need to make some changes to the committee based on people's availability, making sure we have different sets of experts because we're going to be looking at other neurological conditions besides stroke. So we want to make sure we have dementia experts, stuff like that.

We're going to be looking at dementia, delirium, Parkinson's, epilepsy and whatever else comes in. So again, if you know of measures that would fit into those categories please let us know or let the developers know about our call.

And we're closing that call for measures July 13th also and then we're going to send them to you all right after Labor Day to begin that review process. And we're looking at work group calls in mid-September and then our steering committee meeting is October 3rd and 4th. So I'll follow up with you all by email later this summer to assess your availability and everything, but just

keep that in mind. You'll be getting another batch of measures at the end of the summer.

And are there any questions? Okay, that's all the next steps.

MS. JOHNSON: Thank you guys.

CO-CHAIR KNOWLTON: Thank you all. It's a great group, great session.

DR. BURSTIN: Thanks, everybody. Thanks to the Davids.

(Whereupon, the foregoing matter went off the record at 3:11 p.m.)

I

J

O

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Neurology Endorsement
Steering Committee

Before: NQF

Date: 06-21-12

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter